Protein Assembly for a Functional Fibrous Product

Ziyan Zhao

Department of Chemical Engineering and Biotechnology

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Murray Edwards College 2019

I would like to dedicate this thesis to my loving family…

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed limit of 65,000 words and 150 figures as set by the Engineering Degree Committee of the University of Cambridge.

Ziyan Zhao

2019

Acknowledgements

First, I would like to thank my supervisor, Prof Elizabeth AH Hall, for her kind guidance, countless help, warm encouragement, and continuous support throughout the time in pursuit of my PhD degree. Her intelligence and diligence are always the best role model which I wish to follow. Apart from providing me the equipment and hardware to finish my PhD, she also creates such a nice environment and mental support within the group, which helps me go through many hard times. Her fancy ideas, amazing teaching and presentation skills and polished writing styles always inspired and encouraged me. And I am grateful to have this opportunity to be her student throughout the 5 years in Cambridge, which is a valuable memory to me.

I would like to thank Prof Eugene Terentjev from the Department of Physics, University of Cambridge for his guidance and help with the mechanical strength measurements. I would also like to give my gratitude to Dr Matthew Patridge for replacing and fixing the sensor box of the Langmiur Blodgett trough. My thanks also go to Dr Huiyee Chee, Dr David Bailey for the discussion of molecular biology work; Prof Tony Cass and Dr Ljliana Fruk for their suggestions in correcting my thesis; Dr Adrien Fisher and Dr Ljliana Fruk for their suggestions in my first-year viva; and Dr Alex Routh and Dr Qian Sun for the collaboration work and publications.

I would also like to acknowledge Cambridge Trust and China Scholarship Council for their generous financial support.

Many thanks go to all the past and current members in Cambridge Analytical Biotechnology group for their companion, help, and many valuable discussions. Many thanks also to all of my friends outside the research group. They all make my time in Cambridge filled with happiness.

Finally, special thanks to my loving parents and my husband, for their patience, constant support, and most importantly, their unconditional love.

Protein Assembly for a Functional Fibrous Product

Ziyan Zhao

Natural protein-based materials are exploring their new applications from the traditional uses. Structural proteins provide scaffolds, whereas functional proteins carry essential biological activities through millions of biochemical reactions. The idea of implementing functionalities into natural structures could provide manufactured protein products with a better fit for human desire. The development of synthetic biology and molecular assembly methods illustrates possibilities for the production of functional structures in-situ, which provides better connections between the functional partners with the structural scaffolds.

Protein fibres are one of the natural structures which possess a unique shape and superior mechanical properties. Apart from the natural protein fibres, many disease-related peptides self-assembled into amyloid filaments and fibrous structures. Natural globular proteins could also form fibres through the directed assembly by changing their storage conditions or through fusion. Elongated polyglutamine peptides cause many neurodegenerative diseases as they assemble. In this thesis, a polyglutamine peptide (Q77) was fused with functional partners to direct the protein assembly in vitro. The role of the polyglutamine was studied during assembly and after the formation of a self-supportive fibrous product. The extensibility of traditionally size-limited fibrous materials formed by disease-related peptides was tested experimentally for the first time. The resultant fibrous product with embedded functionalities mimics the structure of silk, but the mechanical behaviour of collagen.

Two structurally distinct proteins were chosen as the functional partners for Q77: a monomeric red fluorescent protein (mcRFP), which is relatively small in size and possesses a b-barrel structure, and firefly luciferase (Luc), which is a larger protein with a fragile structure consisting of two mobile domains. Both proteins have been widely used as reporters for intracellular activities with either fluorescence or bioluminescent signal. In this work, the functionalities of both proteins were investigated after Q77 fusion and after assembly towards respective fibrous products. The structural variation of these recombinant proteins resulted in the changes of their functionalities. Finally, a self-supportive fibrous ATP sensor was achieved for the first time with this dual functional protein product.

Publications

Candidate's contribution is an extension of the work in this thesis but not included here.

Sun Q, Du Y, Zhao Z, et al. Functional silver-coated colloidosomes as targeted carriers for small molecules[J]. Langmuir, 2017, 33(15): 3755-3764.

Hall E A H, Chen S, Chun J, et al. A molecular biology approach to protein coupling at a biosensor interface[J]. TrAC Trends in Analytical Chemistry, 2016, 79: 247-256.

Sun Q, Zhao Z, Hall E A H, et al. Metal coated colloidosomes as carriers for an antibiotic[J]. Frontiers in chemistry, 2018, 6.

Contents

List of Figures

List of Tables

Nomenclature

Aβ	amyloid-β
AMP	adenosine monophosphate
APS	ammonium persulfate
ATP	adenosine triphosphate
a.u.	Arbitrary unit
AWI	air-water interface
BCA	bicinchoninic acid assay
BRET	bioluminescent resonance energy transfer
BSA	bovine serum albumin
CD	circular dichroism
D-LH_2	D-luciferin
DNA	deoxyribonucleic acid
DsRed	Discosoma sp. red fluorescent protein
DTT	dithiothreitol
ECM	extracellular matrix
E.coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EHT	electron high tension

fLuc	firefly luciferase
FRET	Förster resonance energy transfer
FT-IR	Fourier-transform infrared spectroscopy
GFP	green fluorescent protein
Gly	glycine
GST	glutathione-S-transferase
HA	hydroxyapatite
His-tag	6 × histidine tag
hMSCs	human bone marrow-derived mesenchymal stem cells
IBs	inclusion bodies
IEP	isoelectric point
IPC	isoelectric point calculator
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	infrared radiation
KAN	kanamycin
kDa	kilo Dalton
LB	Lysogeny broth
LB trough	Langmiur Blodgett trough
LC	liquid condensed
LE	liquid expanded
LR	Luc-mcRFP
LRQ	Luc-mcRFP-Q77
LQ	Luc-Q77

Luc	luciferase
mcRFP	monomeric cysteine mutated red fluorescent protein
mRFP1	monomeric red fluorescent protein 1
MALDI	matrix-assisted laser desorption/ionization
MW	molecular weight
nm	nano-meter
ns	nano-second
OD_{600}	optical density at a wavelength of 600 nm
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCDs	protein conformational disorders
pDNA	plasmid DNA
PMT	photomultiplier
PNAC	protein and nucleic acid chemistry
polyQ	polyglutamine
PVDF	polyvinylidene fluoride
QY	quantum yield
Q77	glutamine 77 repeats
RGD	tripeptide arginine-glycine-aspartic acid
Rh101	rhodamine 101
RhB	rhodamine B
RLU	relative light unit
RNA	ribonucleic acid

SAM	self-assembled monolayer
SDS	sodium dodecyl sulfate
SEM	scanning electron microscope
Ser	serine
TagBFP	a monomeric blue fluorescent protein generated from TagRFP
TEMED	tetramethylethylenediamine
ThT	thioflavin-T
THz	terahertz radiation
Tyr	tyrosine
UHP	ultra-high pure

Chapter 1 Introduction

1.1 Protein Structures: from Single Molecules to Biopolymeric Products

1.1.1 Protein Molecular Structures

Proteins, which come in many different sizes and shapes, constitute a diverse class of biomolecules that underlie most of life's functionalities.[1,2] A single cell contains thousands of proteins, each with its own unique function. Although their structures and functions vary significantly, all proteins are made up of one or more chains of amino acids connected by peptide bonds. The diverse functionalities of proteins are determined by their unique structures, ranging from the primary, secondary, tertiary and quaternary structures. A comprehensive understanding of protein basic structures is critical for productive protein modelling, including structural prediction and de novo design, as well as understanding protein misfolding diseases.[3] The primary structure refers to the linear sequence of amino acids held together by covalent bonds. The cleavage of the primary peptide chain occurs by either chemical hydrolysis or proteases. In addition, certain amino acids (e.g. tryptophan) are able to absorb UV light. Upon absorption, they can transfer an electron to the nearby covalent bonds, causing them to break.[4]

Secondary structures normally refer to the local folded structures that form within a polypeptide due to interactions between atoms of the backbone. The most common types of secondary structures are the α helix and the β pleated sheet. Both structures are caused by hydrogen bonds, which form between the carbonyl (C=O) oxygen atom of one amino acid and the amino hydrogen (N-H) atom of another. In an α helix, for example, the amino hydrogen is 4 amino acids away from the carbonyl oxygen. This pattern of bonding allows the polypeptide chain into a helical structure with each turn of the helix containing 3.6 amino acids. In many structural

proteins such as cytoskeleton, hair and wool, α helices are found to be essential in determining the mechanical properties of these protein materials.[5] Coiled coil is a special helical structure in which 2-7 α-helices are coiled together like the strands of a rope. The most common types of the coiled coil structure are dimers and trimers. Coiled coils could stabilise the proteins in the water-filled environment of the cytoplasm by sandwiching the hydrophobic side chains between the hydrophilic residues.[6] A leucine zipper is an example of dimer coiled coils, which display a periodic repetition of leucine residues at every seventh position over a distance covering eight helical turns. The leucine residues from each helical strand weakly interact with each other, reversibly holding their α-helices together and forming the dimer structure.[7]

β-sheet domains play a key role in forming much stronger protein materials such as muscle tissue, spider silk, and amyloids.[8] In a β pleated sheet, the sheet-like structure is formed by two or more segments of a polypeptide chain lining up next to each other. The hydrogen bonds form between carbonyl and amino groups of each segment. The strands of a β pleated sheet can be parallel, pointing in the same direction (meaning that their N- and C-termini match up), or antiparallel, pointing in opposite directions (meaning that the C-terminus of one strand is positioned next to the N-terminus of the other).

Apart from the carbonyl hydrogen bonds, other forces are also involved in stabilising protein secondary structures, such as the n→π* interactions and C5 hydrogen bonds in the main chain.[3] The n→π* interaction denotes the force between adjacent carbonyl groups in the backbone. It contributes mostly to the α helix and other helical confirmations.[9] The C5 hydrogen bonds occurs between the amide proton in β strands and its own carbonyl oxygen. These C5 hydrogen bonds are shown to stabilise amyloid formation, which is implicated in many neurodegenerative diseases.[10]

The tertiary structure is caused by interactions between the R groups of the amino acids that make up the protein. R group interactions that contribute to tertiary structure can be hydrogen bonding, ionic bonding, dipole-dipole interactions, hydrophobic interactions, π-π interactions, London dispersion forces (non-covalent bond) and the disulphide bond (covalent bond). Figure 1.1 compares the formation of basic secondary and tertiary structures.[11–13]

Monomeric proteins are made up of a single polypeptide chain and possess only the above three levels of structures. However, some proteins are made up of multiple polypeptide chains, also known as subunits. When these subunits come together, they give the protein its quaternary structure. This structure is omitted as it is out of the scope of this thesis.

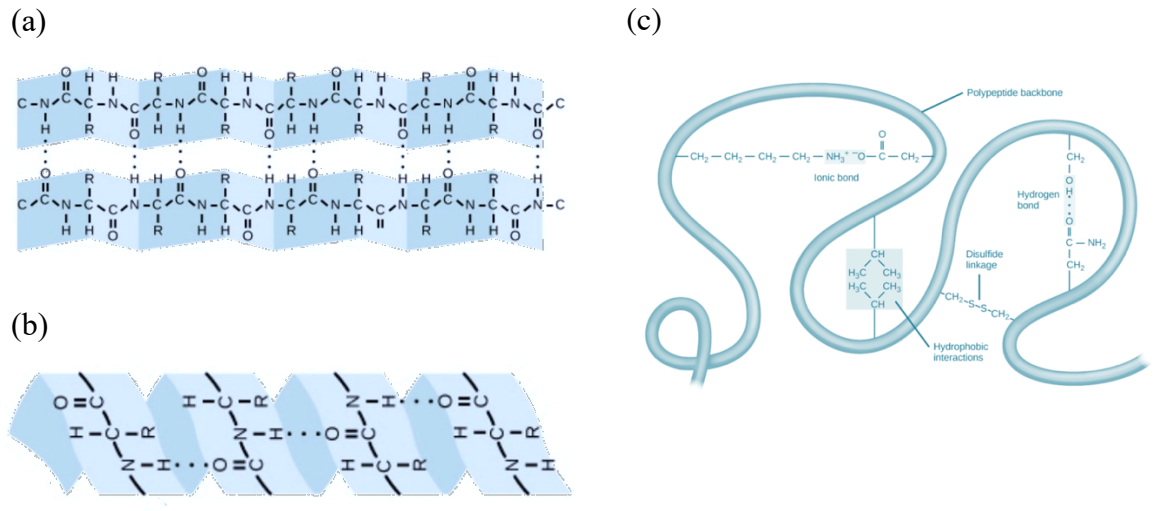


Figure 1.1– Basic structures of proteins: (a) α helices, (b) β pleated sheet, and (c) tertiary structures. (reproduced from Openstax,[14] access for free at https://openstax.org/books/microbiology/pages/1-introduction)

1.1.2 Protein Aggregation and Self-assembly

Proteins may undergo structural changes naturally or under external stresses, which result in protein denaturation. During the denaturation process, proteins lose their quaternary, tertiary, and secondary structures which is present in their native state. Modern techniques have been widely used to study and monitor the structural changes such as circular dichroism, infrared/X-ray spectroscopy, optical and electronic microscopic methods with or without labelling. Natural protein denaturation may be a consequence of cell death or an indication of certain diseases.[15] Recent research shows diverse diseases arise from protein misfolding and are categorised together under the name of protein conformational disorders (PCDs).[16] Structural changes in these cases may be induced by pathological chaperone (apolipoprotein E, amyloid P component, protein X), modification on the environmental conditions (pH, oxidative stress,

metal ions), and also the original gene mutations which result in a different coded protein and thus a misfolded one from the native state.[17] Among these pathways, changing the external conditions have also been utilised to purposely modify native protein structures for aggregated products in vitro. The external stress applied to proteins in vitro could be expanded to conditions such as high voltages, radiation, heat, and denaturing chemicals. These aggregated products through misfolding can be utilised in areas such as drug delivery,[18] wound healing, and biosensing.[19]

While protein misfolding may be designed for a variety of applications, nature also has its own way to create aggregated structures, which is achieved through self-assembly. Self-assembly is a process in which a disordered system of pre-existing components forms an organised structure or pattern as a consequence of specific, local interactions among the components themselves, without external direction.[20] It is a multidisciplinary process which is inspired by bio-systems, originated from chemistry, follows the physical and chemical principles, and can be applied to the functional materials. Self-assembly may occur on designed nanomaterials, biological and artificial molecules, or surfaces. It is based on intermolecular interactions such as van der Waals interaction, hydrogen bonding, hydrophobic interactions, dipole-dipole interactions, ion paring and π-π stacking.[21]

Many biological macromolecules already possess organised structures at their molecular scales. These organised structures are generally connected by covalent bonds through polymerisation, which is not an assembly process. A single-stranded DNA, for example, is built up by 4 different nucleic acids through polymerisation. The dimerization of the single stranded DNA to the double-helical structure through hydrogen bonds, however, happens without external forces, which is a self-assembly process. Furthermore, the single-stranded DNA could be built up by the more recent technique, termed DNA origami, with the help of synthetic oligomers or the staple strands to create various shapes at nanoscale levels.[22] Similarly, other biological building blocks, such as oligonucleotides, oligosaccharides, phospholipids, peptides or proteins, may self-assemble into ordered three-dimensional, multi-component structures at their supramolecular scales, forming natural bio-polymeric scaffolds.[23] These assembled bio-polymeric materials are widely used in many human activity related fields, for example, tissue engineering, regenerative medicine, and drug delivery.[24] Among

these structures, protein scaffolds possess many superior properties. In nature, protein scaffolds have several major advantages over other biopolymers such as the superior mechanical properties, large varieties of biological functions, and easy modifications due to the high number of accessible functional groups on protein backbones, depending on the amino acid composition. However, the methods for the design of self-assembling proteins have developed more slowly compared to the development in the fields of DNA, RNA, and synthetic peptides, as a result of the large sizes of proteins, chemical heterogeneity, and the complexity of protein–protein interactions. In spite of these, there are several approaches proposed for the design of artificial self-assembled protein, which includes protein fusion, chemical conjugation, and computational-aided de novo design. A large amount of protein complexes analysis suggested that most assembly steps can be classified into three basic types: dimerization, cyclization, and heteromeric subunit addition.[25]

Self-assembly of proteins often results in aggregates with high stability, insolubility and resistance to degradation. Some proteins are designed to go through the self-assembly process for different applications. For instance, the 24 subunits of ferritin are held together by non-covalent interactions, arranged in an icosahedral cage with a central cavity. This cavity is used to store and enucleate iron in nature, but it can also be disassembled and reassembled for the encapsulation of drugs or dyes.[26] Recent research also discovered that two proteins, 53BP1 and RIF1, assemble into a circular scaffold to protect genome integrity.[27] There are also a variety of proteins that self-assemble to make up the outer shell of bacteria or viral capsids in their natural status.[25]

Other proteins require site mutations, chemical modifications, or certain environmental conditions such as pH, ionic strength, consulates, and temperature, to trigger a conformational change, which result in protein or peptide assembly into different structures such as crystals, filaments, gels, across a range of length scales from nm to μm (Figure 1.2). [25,28,29]

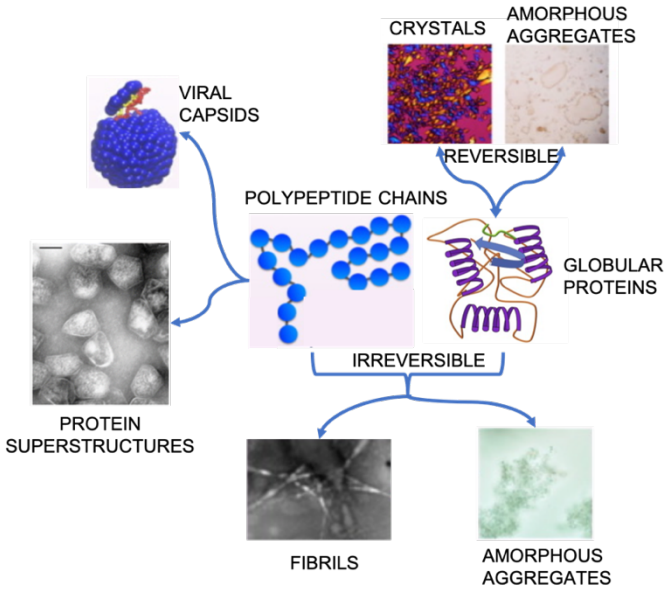


Figure 1.2– Peptide and globular protein self-assembly into different structures. (Reproduced from McManus[25], Fan[28], and Shen[29] et al)

1.1.3 Protein Self-assembly for Fibre Formation

Fibres are one of the most interesting and important structures found in nature, whose width is negligible in comparison with their length. Biological fibers are natural materials that possess the property of high mechanical strength, high elasticity, high thermal stability, and biocompatibility.[30] Protein fibers are either naturally occurring to fulfill the structural roles in living organisms, or through peptide and globular protein misfolding which leads to severe human diseases. The formation of such fibrous peptide or protein materials has attracted many research interests, not only for the study of diseases but also for their broad applications in bionanotechnology as a new type of material, ranging from nano-optoelectronics, drug delivery, to biosensing and bioremediation.[23,31–33]

Natural globular proteins that have the potential to form fibres normally go through a step of either partial structural unfolding or local conformational change, leading to the molten globule state of protein structure (Figure 1.3). The molten globule is a protein denaturation intermediate with little detectable tertiary structure and increased solvent-exposed hydrophobic surface area comparing to the native state. These steps will expose certain sites within the protein and trigger

the assembly.[34] Insoluble fibres with high stability will subsequently form from the soluble protein solution after a nucleation-dependent polymerisation process.[35] The resultant fibril assemblies usually have the predominantly hydrogen bonded β-sheet structure. The initial discovery of globular protein assembled into fibrous structure was caused by the mutation of the protein, which further leads to the unfolding process and conversion into amyloid fibril.[36] As is discussed in the previous section, such mutation and the consequent fibril aggregation leads to the protein conformational disease. The assembly of these proteins into fibres can be a spontaneous self-assembly after the formation of the molten globules.

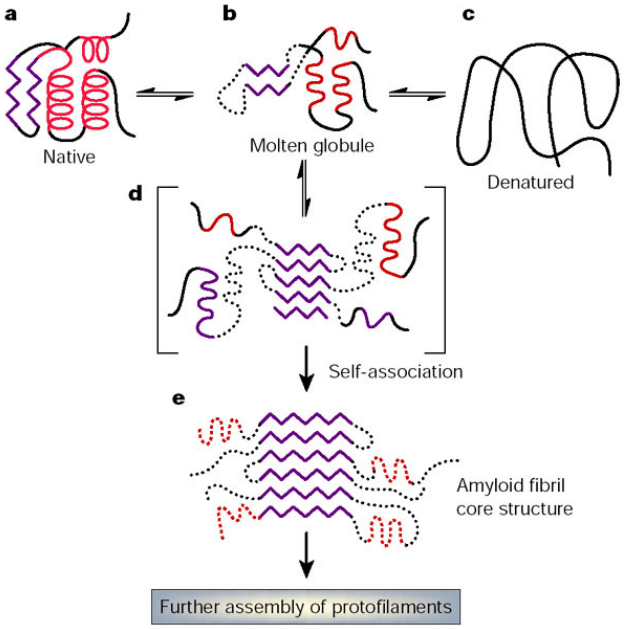


Figure 1.3– Protein conformational changes during self-assembly for fibre formation. (Reproduced from Dumoulin et al)[36]

The most widely studied induction conditions for protein assembly are heating and acidification. It was previously reported that natural globular proteins form fibril aggregates at a pH far away from the isoelectric point (IEP) under heat denaturation; whereas at a pH near the IEP, spherical aggregates were observed with high temperature treatment.[34] Michelle et al further demonstrated that the globular and protofibrils formed as intermediates before the mature fibrils.[37] Domigan et. al[38] had a detailed study on the fibrillation conditions of insulin, showing that high temperature, relatively low pH, and high protein concentration resulted in an optimal fibre structure in all the three dimensions, as well as fast kinetics in fibre formation.

The other protein known to form fibrillar structures is α-Lactalbumin, which is very attractive for studies of partially folded conformations. The protein adopts the classic molten globule state at either acidic pH or evaluated temperature, further aging conditions give rise to the amyloid fibrous structure.[39] Later research suggests an alternative pathway with calcium induced self-assembly of partially hydrolyzed α-Lactalbumin.[40] Recent research also showed the ability of nanoparticles to seed protein fibre growth, together with the traditional induction treating conditions. Proteins such as lysozyme, β-lactoglobulin, and bovine serum albumin have been used as models for investigating the effect of nanoparticles on protein fibre formation.[41] Even though some modified nanoparticles such as cholesterol modified nanogel and negatively charged gold nanoparticles are found to refold denatured proteins and inhibit fibril formation, some in vitro experiments indicate that pure nanoparticles, for example TiO_2 and pure gold nanoparticles, accelerate fibril aggregation.[42] The protein-nanoparticle complex is usually formed through interactions such as solvation forces, hydrogen bonds, van der Waals forces which result in a conformational change in protein structure.[43] The increase in local concentration of nanoparticle complexed protein is explained as the reason for the faster fibre formation.[41–43] Such in-situ growing methods provide a simple and facile route in producing biological fibril materials.

Various short peptide sequences, such as amyloid-β (Aβ), huntingtin protein, prion protein and polyglutamine peptides (polyQ) also form the disease-related fibril structures. These amyloid fibrils have been examined to exhibit superior mechanical properties mimicking silk.[44] However, these properties are limited to the stiffness and toughness, tested through AFM due to their nanoscale sizes; and the molecular origin of these properties are contributed to the largely formed β-sheet structures. The extensibility of amyloid fibrils, however, is only estimated through simulation results.[45]

De novo design of fibrils made of short peptides has shown great success. In one study, the assembly of the coiled coil structure folded from the designed peptide sequence is triggered by evaluated temperature;[46] whereas in another research, the designed peptide assembled into different structures at different pH.[47] Recent advances also successfully design self-assembling helical filaments from monomeric proteins.[48] Chemical modification opens up a new route for protein assembly in their native folds. Kashiwagi et al reported the polymerisation of DNA-

modified chaperon into nanotube structure with controllable length, which is a major breakthrough in designing new drug delivery systems.

1.1.3.1 Self-assembled Protein Fibres for Electrochemical Biosensing

The aggregates in the fibrous form are an excellent physical support to organize inorganic materials, organic materials and even functional enzymes. Apart from the general application areas such as drug delivery and tissue engineering, fibrous aggregates in biosensing also attract many research interests. The integration of assembled protein fibres with conductive materials provides these materials with the electron transfer property which is essential for designing the electrochemical biosensors. An electrochemical biosensor generally consists of a biorecognition component which specifically captures the analyte; and a signal transducer which transfer the electrons generated from the biochemical reaction to a conductive material. The signal is then reported by either potentiometry, impedance or conductivity. A suitable combination of the transducer and the biorecognition component is essential for an effective biosensor. Nanomaterials or self-assembled monolayer (SAM) modified electrodes, for example, may guide certain redox enzymes in a desired orientation to achieve an efficient electron transfer.[49–51] A specific biorecognition component such as DNA, antibody, or redox enzymes for a certain biomarker, on the other hand, improves the specificity and sensitivity of the sensor.[52] Recent advances in designing electrochemical biosensors have led to the development of the *in vivo* biosensing-drug delivery system for diabetes monitoring and treatment.[53]

On one hand, the assembled protein fibres may act as the transducer to transfer the electrical signal. Even though nanoparticles added in protein solutions directed the assembly of monomeric protein molecules, they did not generate any conductivity in the assembled product due to the large distances between the nanoparticles. The readily aggregated protein fibres, on the contrary, may be decorated with the nanoparticles for the conductivity purposes. For instance, covalently linked silver nanoparticles have been decorated onto the amyloidogenic protein NM Sup35 (the N-terminal and middle region of yeast prion Sup35) for the fabrication of a conductive nanowire.[54] Similar idea has been used in designing a novel non-enzymatic H_2O_2 electrochemical sensor. In this case, the protein fibre self-assembled from a specifically

designed peptide templated the synthesis of silver nanowires. Further assembly of the silver nanowires on graphene nanosheets enhanced the performance for electrochemical determination of H_2O_2.[55] A sandwiched metal-peptide-metal structure was also fabricated by decorating gold nanoparticles onto silver-filled diphenylalanine nanotubes.[56] Apart from these metal nanoparticles, amyloid fibres can also be decorated with conjugated polymers. Acidification and heating induced lysozyme fibril has been polymerised with a conducting polymer, polyanaline, which has a good conductivity and various oxidation states.[57] Similar methods with sulfonated polythiophene coated insulin amyloid fibres have been used to fabricate microfiber electrochemical transistors.[58] The successful incorporation of heme molecules onto a cytochrome b attached amyloid fibril has provided a pathway mimicking the natural long-distance electron transfer.[59] Such templates showed a promising possibility for the protein fibrous structure as a conductive material in electrochemical biosensing applications. Furthermore, the integration of self-assembled peptide nanostructures in micro-fabrication processes shows great application in the development of the highly sensitive "lab-on-a-chip" devices.[60–63] Recent advances also show that fibrils made of peptides substitute with aromatic groups present efficient electron delocalization, and thus provides the material with electron transfer property.[64,65] Such discovery not only facilitates new inventions in biosensing but may also be utilised in solar-cell applications.[66] Self-assembled monolayer of peptides, for instance, was used to modify the surface of the gold electrode. The electrical properties of the electrode were found to depend on minor changes of the electrode surface, due to peptide phosphorylation by different kinase.[67] Therefore, a highly sensitive and selective detection method for kinase activity was developed. In another example, self-assembled peptide scaffolds have been used to create light harvesting supramolecular structure that mimic the natural photo synthesis process.[68]

On the other hand, protein fibres have been utilised as biorecognition components to detect certain analyte. An amyloid fibril-functionalised screen-printed carbon electrode has been devised for the determination of metal ions. A reduced current was observed due to the binding of the metal ion to the amyloid fibril, which inhibited the redox reaction of amyloid fibril.[69] In another study, protein nanofibers of β-lactoglobulin attached to the surface of graphene has been designed to monitor the present of porcine gastric enzyme. It is observed that the protein

nanofibers could effectively prevent the aggregation of graphene nanosheets. The enzymatic degradation of the protein nanofibers change the conductivity of the graphene film.[70] These conducting protein fibrous structures decorated either by nanomaterials or organic molecules are excellent materials for applications, due to their conductivity, high mechanical strength, high surface attachment to substrates, high biocompatibility comparing to the classical polymeric systems, and their resistance to thermal and chemical manipulations.

1.1.4 Recombinant Protein Fibres

Nature creates different structures for mechanical functions, through the process of self-assembly. Natural structures such as shells, trees, skeletons, nests act as physical supports. The incorporation of natural active components granted most of the biological activities, which have been tuned through millions of years of evolution.[71] The extraction of natural biomaterials shows great application in many human activities. While the detailed study of the extracted component often requires the purity of the material, further modifications or different components recombination usually allow them to be utilised in more realistic settings. In such cases, synthetic biology became an important method for the production of recombinant proteins with multiple functions. It opens up a facile route to embed functionalities more evenly into protein structures *in vitro*, avoiding the later multiple processes of chemical treatments. As is discussed earlier, both the denatured globular proteins and many disease related peptides have the potential to self-assemble into fibrous structures. The fused protein assembly usually involves two different naturally occurring domains fused into a single protein chain. Oligomeric domains that began or ended in an α-helix is a common complex that is used in this method since an unbroken α-helix running between two domains meets the required directionality and rigidity for self-assembly. This strategy was first used for designing a 12-subunit molecular cage with a tetrahedral shape and protein filaments and later an asymmetry matching method was introduced which simplify the design process.

In this case, the peptides are much easier to control and are shorter without complex tertiary structures comparing to the denatured globular proteins. Therefore, they are more suitable as the fusion partner with functional proteins. Fusion of peptides with functional partners have seen many successes. For example, fluorescent proteins and enzymes such as Glutathione-S-

transferase (GST), barnase and carbonic anhydrase, have been fused with [URE3]. [URE3] encodes the self-propagating infectious protein (prion) of *Saccharomyces cerevisiae*, into fibrils in their active forms.[72] The growing condition in this case depends highly on the fibre formation part in the construct. The fibre formation parts utilise natural fibrous proteins which have a large potential to self-assemble in mild conditions. The [URE3] protein in a very low concentration of 0.1-1.0 mg/mL, for example, can be grown at 4 °C, physiological pH for a minimum of 16 hours to get fibril aggregates.[72] Such method also produces functional fibrous materials in a moderate condition without destroying the structure of the functional protein.

1.1.4.1 Structural Elements for the Protein Fibre

Among the disease-related fibril forming peptides, polyglutamine (polyQ) attracts many research attentions due to their special amino acid composition and folding structures. Polyglutamine (polyQ) is a peptide sequence with single amino acid repeats in eukaryotic proteomes, which has been reported to self-assemble to amyloid fibrils both in vitro and in vivo when the glutamine repeats exceed a certain threshold. It can be found in a variety of protein families which do not appear to be related such as the huntingtin protein and the spinocerebellar ataxias.[73] Polyglutamine diseases are a group of neurodegenerative disorders caused by the aggregation of the abnormally elongated polygultamine repeats. While 17-20 glutamines are typically found in normal non-mutated proteins, expression of huntingtin protein, for example, with an expanded chain of greater than 35 glutamines are associated with Huntington's disease. The aggregation of polyglutamine (polyQ) has been studied intensively in vitro since its discovery; whereas the widely accepted cellular toxicity from the soluble form of the expended polyQ oligomers[74,75] has been challenged recently as it is proved that the large intracellular aggregates produce long term toxicity *in vivo*.[76]

Rapid aggregation in vitro towards amyloid fibrils is typically observed under mild conditions for Q35 peptides and larger, while for repeat lengths smaller than Q25, aggregates can only be obtained under harsh conditions and stresses, such as chemical denaturation or heat.[77–79] A method developed by Streets et al. shows that the fibril size distribution does not consistently follow the same kinetic evolution as β-sheet content with increased concentration, which suggests a complex kinetics with some degree of fibril bundling. [80] Even though the secondary

structure[73] and the unfolding process[81] of the polyQ domain are controversial and suspected to be protein partner dependant, the cross-β-sheet is widely accepted as a characteristic structure of the amyloid fibril product. Scherzinger et al. studied the non-aggregation property of a GST fused pathogenically expanded huntingtin protein, which further indicates the aggregation property of the polyQ domain is largely protein dependant.[82] Similarly, research done by Jiang et. al proves the different toxicity effect of the huntingtin protein through fusion with a variety of fluorescent proteins.[83]

Native protein isolated from the cell can be refolded in vitro from solubilized inclusion bodies (IBs) either by dilution, dialysis or on-column refolding methods.[84] It has been shown that formation of IBs is usually accompanied with the formation of significant new β-sheet structures, stabilized by inter-chain hydrogen bonds, causing tightly packed, extended intermolecular β-sheets.[85] A typical example of this structural transition peptide is Amyloid-β, whose accumulation is proposed to be an early event in the pathogenesis of Alzheimer's disease. This transition can be due to facial amphiphilicity, causing hydrophilic domains to be presented at the molecule/water interface, while protecting lipophilic facial domains from interaction with the water. However, other mechanisms can also result in aggregating structures. For example, in polyglutamine (polyQ) diseases, there are several models proposed, with the presence of either α–helical or β–sheet structures; Buchanan et al, for example proposed the structure for polyQ to consist of stacked β-hairpins.[86] However, glutamine is interesting because it is a polar amino acid and, despite the absence of hydrophobic residues. Rhys and Dougan[87] have shown that hydrogen bonding between the side chain and backbone, is associated with the collapse of the polyQ into an insoluble structure due to internal hydrogen bonds. This takes preference over hydrogen bonding with water for chain lengths >15 residues, so that water becomes a poor solvent, and the conformation attains a high mechanical stability. Side-chain – side-chain hydrogen bonds then add to the stabilisation energy resulting in aggregates with cross β-sheet architecture.

1.1.4.2 Functional Elements for the Protein Fibre

Red fluorescent protein was chosen as the initial functional exemplar building block due to its tripeptide chromophore, formed during the folding process and susceptible to configuration

distortion during folding. The ease of measuring the absorbance and fluorescence properties of this protein make it a good fusion partner to explore the potential for fibrous functional proteins. Similarly, the enzyme luciferase is a first exemplar for a fibre biosensor. It is a protein which is easily denatured due to its susceptibility to unfolding. It thus provides a challenging fusion partner to achieve fibre functionality.

1.1.4.2.1 Red fluorescent proteins

Fluorescent proteins have been widely used as reporters to investigate a variety of cellular activities since the first discovery of a green fluorescent protein (GFP) in the crystal jellyfish *Aequorea Victoria* in the 1960s by Osamu Shimomura.[88] The mystery of the fluorescent protein family gradually unveiled after a GFP-like protein, DsRed from *Discosoma sp*., was reported.[89] This DsRed protein extended the coverage of the colour spectrum that GFP and its derivatives did not previously reach. A number of red fluorescent proteins that emit orange, red, far-red fluorescence have been isolated and engineered through molecular evolution. Since then, the structural basis for the red fluorescence, and the mechanism of the chromophore formation draws the attention by various researchers.

Similar to GFP, DsRed folded into a β-barrel structure preventing the chromophore from quenching by the solvent. DsRed possess the Ser-Tyr-Gly structure, a chromophore forming tripeptide, within the barrel (Figure 1.4 a). Many of the currently know RFPs share the DsRed-like chromophore, which has the same structure with the GFP chromophore with an additional N-acrylimine (O=C-N=C) group. Many post-translational steps are required for the maturation of the chromophore including the initial cyclisation and dehydration/oxidation, which results in the conjugated π-systems with different lengths and emits different wavelength over the spectrum. The formation of a green and blue chromophore during the maturation was confirmed by the absorption spectrum since the discovery of DsRed.[89] Different hypotheses, however, were developed for the mechanism of red chromophore maturation over the years: the branched pathway, the "via-GFP" and the "via-TagBFP" model. All models agree with the initial cyclisation step, which involves the reaction between the carbonyl carbon of Ser65 and the amino nitrogen of Gly67, forming a 5-membered ring. The first model proposed was the "via-GFP" pathway,[90] which suggested the formation of the red chromophore via the green

intermediate (Figure 1.4 f). The green intermediate is formed from the O_2-mediated oxidation of the Cα-Cβ bond of the Tyr66 chromophore (Figure 1.4 b→e→f). The consequent formation of the N-acrylimine group leads to the final maturation of the red chromophore (Figure 1.4 f→g). Later on, it was proposed that the true intermediate is the protonated green intermediate.[91] The anionic form proposed in the "via-GFP" pathway appears to be a dead-end product. All oxidation steps occur before the formation of the blue intermediate. This model, however, is at variance with experimental observations that there is no blue to red chromophore transition in the absence of oxygen after the formation of the blue form, as well as some quantum mechanical calculations, revealing the cationic structure of the blue form to be unlikely. Therefore, the "via-TagBFP" pathway is shown to be the most sensible explanation.[92] It is proposed that the red chromophore could be transited from both the blue (Figure 1.4 b→c→d) and the green intermediate (Figure 1.4 b→e→f). The existence of the green absorbance in the matured DsRed was explained by the reversible reaction of N-acrylimine formation (alkane oxidation, i.e. dehydrogenation, Figure 1.4 f→g) and dissociation (alkane reduction, Figure 1.4 d→g); whereas the blue intermediate is almost depleted after the red chromophore maturation.

Among all the DsRed variants, the development of a monomeric red fluorescent protein (mRFP1) largely reduced the maturation time (from 10 hours to 1 hour) and simplifies the protein structure for fusion. The mRFP1 maintains the red fluorescence with the quantum yield being 1/3 of the wild type DsRed. It also retains partial blue absorbance after the maturation of the red chromophore.[93]

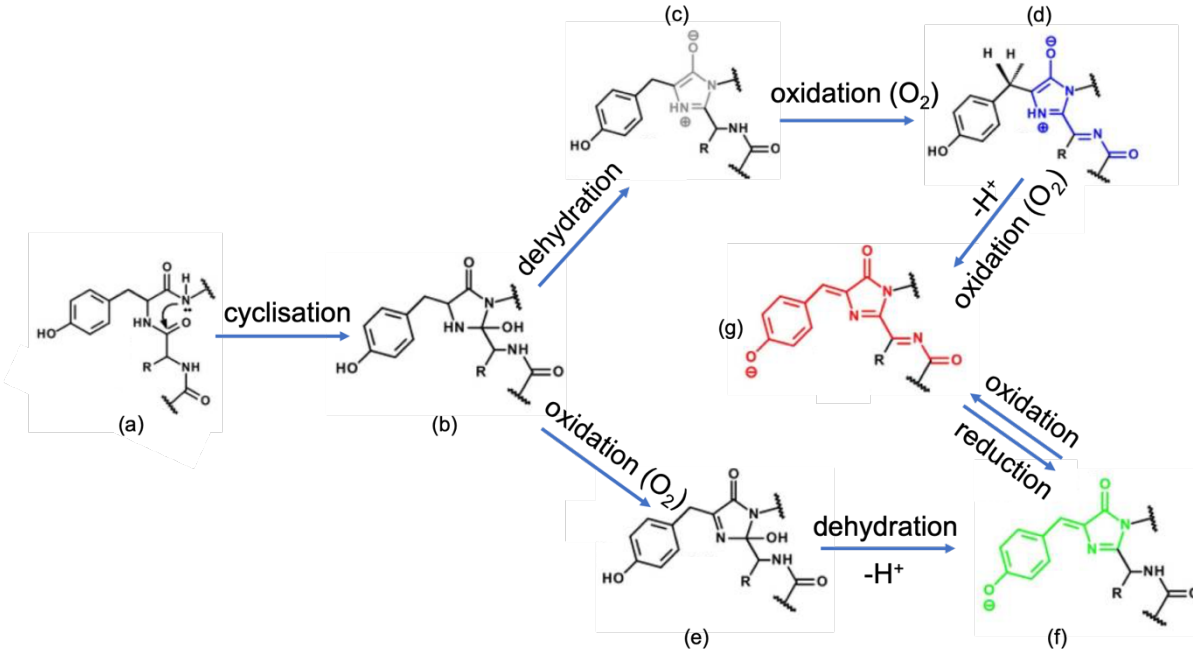


Figure 1.4– Chromophore maturation pathway of DsRed described by the "via-TagBFP" model. (Reproduced from Subach et al)[92]

1.1.4.2.2 Firefly luciferase

Firefly luciferase (FLuc) is another common reporter protein with catalytic activity, which is widely used by generic fusion to study cellular event, due to its bioluminescent property with easy quantifications. It is also used for *in vivo* bioluminescent imaging, to simultaneously visualise and monitor the activities of mice. Luciferase has different variants from different organisms, e.g. firefly, beetles, marine organisms, and many bacteria. Different luciferase variants use different substrate and emit light at different wavelength vary from green to red (Table 1.1). Bruce et al reported a dual-colour assay using two variants of luciferase, which could selectively react with their respective substrates with the emission of well-separated spectra of both green and red light.[94]

Table 1.1 – Common luciferases, their luciferin substrates, and the emission wavelengths. (Reproduced from Kirkpatrick et al[95] and Widder et al[96])

Luciferase	Luciferin substrates	Emission Wavelengths (nm)
Firefly luciferase	D-luciferin	560
Click Beetle luciferase	D-luciferin	613
Bacterial luciferase	Tetradecanal	490
Renilla luciferase	Coelenterazine/$FMNH_2$	480
Oplophorus luciferase	Coelenterazine/Ca^{2+}	460
Gaussia luciferase	Coelenterazine	470

Among all the variants of luciferase, Fluc is the most widely studied one. Firefly luciferase catalyses the oxidation of D-luciferin (D-LH_2) in a two-step reaction, which results in the emission of light:

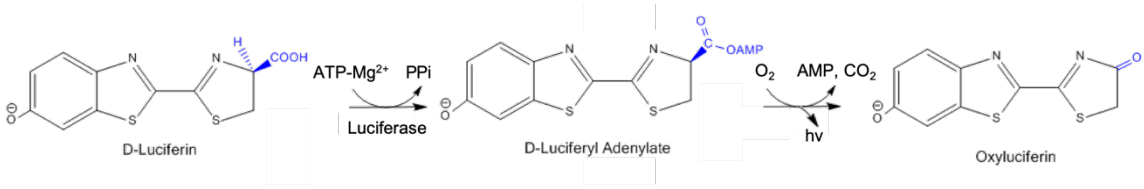


Figure 1.5– Two-steps oxidation of D-LH_2 catalysed by firefly luciferase. (Reproduced from Baldwin)[97]

The exited state of the final product, oxyluciferin, formed as an intermediate when D-luciferyl adenylate is oxidised by the environmental oxygen. The excited oxyluciferin tautomerizes between the keto and enol form.[97] The light emitted is determined by the form of the excited oxyluciferin as it returns to the ground state. It is suggested that red light is emitted from the keto form, while green light is emitted from the enol form.[98] Recent study also examines the

dependence of microenvironments on the bioluminescence colour.[99] The Fluc from *Photinus pyralis*, for example, emits the higher energy green light at ambient conditions and physical pH. However, the colour exhibits a large red shift at low pH and under destabilizing conditions.[100] The Fluc used in this thesis is the P. pyralis luciferase with 5-point mutations, which provide the luciferase with improved thermostability and no red shift at acidic pH.[101] A special fact for luciferase is that it emits light that flashes and disappears within a few seconds. This is because that the final product, oxyluciferin, acts as a competitive inhibitor to the enzyme.

Luciferase is one of the most interesting enzymes to study, due to its capability in engineering wavelength shifts, through either mutageneric engineering, or changes in environmental pHs. Branchini et al produced a red-shifted variant with emission wavelength at 615 nm with improved kinetics.[102] Due to the requirement of ATP, luciferase was developed and modified for many ATP assays, which could further indicate the cell viability. Fluc-based bioluminescent ATP assays display detection limits down to 10^{-17} mol.[96] The ATP concentration in living cells are generally saturating for Fluc, preventing it to be used in the intracellular ATP assays. The application of FLuc in intracellular pH sensing is also well-established.[103] In addition, the fusion tag technology also allows an ATP sensing on immobilised platform. Golnaz et al studied the fusion of FLuc with mCherry, for the development of a BRET-linked ATP assay.[104]

1.2 Natural Protein Fibre Scaffolds

1.2.1 Fibrous Proteins

While a variety of peptides and denatured globular proteins go through the self-assembly processes for the production of protein scaffolds being used in many biomedical applications, scleroproteins, or fibrous proteins, play a key role in forming protein scaffolds in nature, which supports many essential biological functions with even more superior mechanical properties. Scleroproteins are one of the major categories of natural proteins, alongside globular, membrane and disordered proteins.[105] Natural fibrous proteins can be divided into two categories. The first category is the inextensible collagen fibres. Collagen is the most abundant protein in human body that provides strength and structural stability to tissues such as cell surfaces, blood vessels, tendons, cartilage, skin, hair, and bone, depending on the type of the collagen.[106] It consists of amino acids bound together to form a triple helix of elongated fibril structure. All of the 30 types of collagen identified contain at least one triple helix.[107] Collagen also serves as a major component of the endomysium in muscles. The muscle tissue is formed by filaments of actin and myosin that slide past one another.[108] It produces force and motion for many living organisms. The second category of fibrous proteins is the highly extensible fibres such as the elastic fibres.[109] Elastic fibres are present as rope-like structures in ligaments, in the media of arteries and skin. They mainly consist of amorphous elastin core and microfibrils. Functioning together in connective tissues with collagen, elastin provides elasticity to the tissue. Figure 1.6 shows the association of collagen and elastic fibre when forming the bovine aorta and compares the different composition and structure of collagen and elastic fibres. As is shown, each collagen fibrils contain three stranded helical segments of similar structure. These fibrils are bundled together, forming the collagen fibre with high stiffness and low extensibility.[110] In comparison, the filament structure of mature elastin is formed by a much complicated packing arrangement of tropoelastin subunits. Recent study suggested that elastin aggregates self-assembled by tropoelastin inside the cell transfer to growing elastic fibres in the extracellular matrix.[111] This transfer process is directed by different microfibrils such as fibrillin and fibulins.[112,113] The elasticity of elastic fibres is resulted from the hydrophobic domain of elastin, which is a compact amorphous structure with

distorted β-strands, fluctuating turns, buried hydrophobic residues, and main-chain polar atoms that form hydrogen bonds with water.[114]

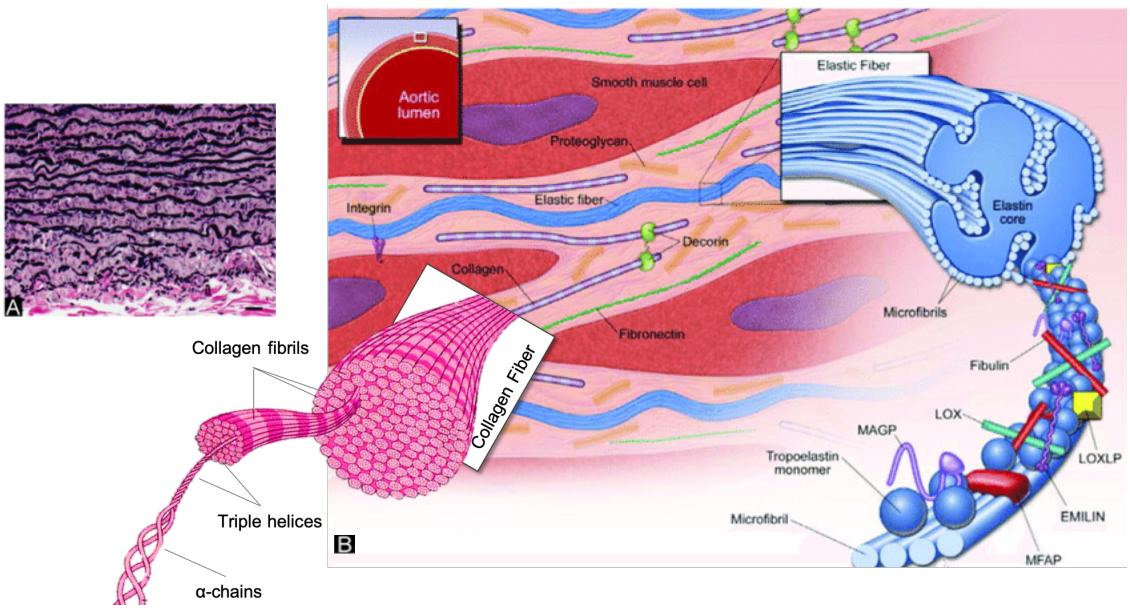


Figure 1.6– (A) Lignt microscopical staining of bovine aorta showing elastic fibres in black and collagen in pink; and (B) structural illustration of the aortic wall. (Reproduced from Daamen[115], Yassine[116] and Daniels[117] et al)

Similarly, α-keratin fibres, together with fibroin fibres such as silks are all highly extensible fibrous proteins belonging to the second category.[118] Apart from these well-known fibrous proteins, reflectin attracted considerable research attentions over the years, due to its unique photonic and remarkable assembly properties. It is an important protein family providing structural colours in cephalopods.[119] Reflectin is widely used to form structures with different shapes[120] and to produce various protein-based optical devices[121] by *in vitro* self-assembly. The latest discovered natural silk-like material is hagfish slime, a cytoskeletal biopolymer. It is a unique predator defence material containing not only slippery mucins, but also threads in a condensed coiled state, known as “skeins”. Skeins possess a remarkable structure wherein a long filament is packed in canonical loops into a spheroid. When mixed with the surrounding water, the fibre unravels from the skein rapidly and forms a fibrous network.[122] Even though a keratin-like component is well studied in the slime thread,[123,124] very little is known about the mucin component.[125]

Among these fibrous proteins, silk possess unique properties of both high toughness and high elasticity. It is an interesting material to study as it covers a wide range of proteins from different species with different structures. Silk has been widely used for various applications such as enzyme immobilisation substrate, in tissue engineering and wound healing.[126–129] Silks are spun into fibres by some Lepidoptera larvae such as silkworms, spiders, scorpions, mites and flies.[130] When insects are required to generate fibers, the precursor proteins will be accumulated in their glands to obtain a viscosity about 3.5 million times that of water. Such high viscosity allows the insects to form continuous fibres by expelling a droplet of the protein onto a substrate and then pulling the solution away from the substrate.[131,132] A typical silk fibre produced by *Bombyx mori* silkworm includes three basic components: the external protein coat, the filament and the sericin. There are two filaments in the fibre which contains the main protein fibroin that is glued together by the protein sericin. For each fibroin, it can be degummed into β-sheet crystalline domain, which is oriented along the fibre axis, and amorphous domain, shown in Figure 1.7.[133,134] Similar to the elastin fibres, these amorphous domain is the main uptake of strain resulted from the unfolding of the polypeptide chains and their orientation by the direction of extension.[118] The fibroin consists of repeated sequence of six residues of $(Gly\text{-}Ala\text{-}Gly\text{-}Ala\text{-}Gly\text{-}Ser)_n$ repeats with 3 protein components: High-chain (350 kDa), Low-chain (26 kDa) and glycoprotein P25 (30 kDa).[135] These structures provide silk fibres with high mechanical strength and low water solubility, which makes them suitable for essential applications such as drug delivery and sensing.

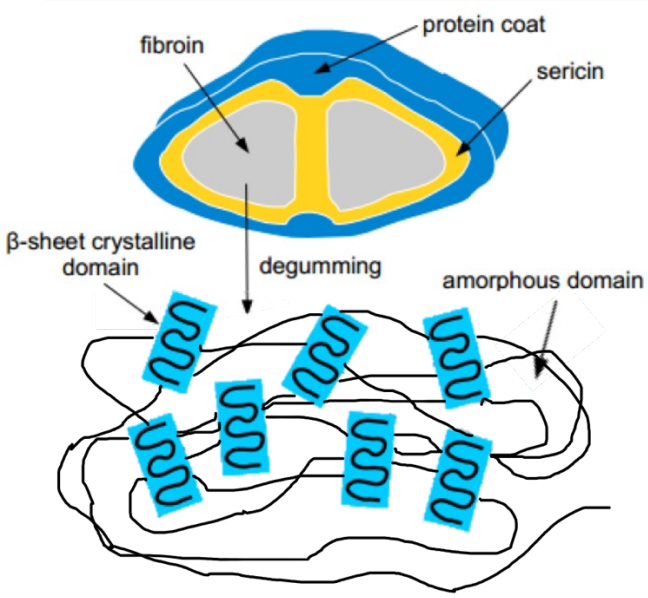


Figure 1.7– Basic components of silk fibre produced by silkworms. (Reproduced from Jao et al)[136]

Apart from the Silkworm cocoons, spider webs and honeybee silk are also the best-known silks being used in manufacturing and medical applications. Similar to the silkworm silk, spider silks consist of internal regions of highly repetitive amino acid "motifs". These motifs are combined into larger repetitive units. Spider silk proteins are rich in β-sheet structures which are aligned the fibre axis.[137] The SEM images of spider silk and silkworm silk are compared in Figure 1.8.[138,139] It is shown that the spider silk consists of layered structure, whereas the silkworm silk is composed of bundled fibrils. In contrast, honeybee silk is helical dominant structure with small β-sheet domains. The honeybee silk is encoded by four small genes (AmelF1-4) with low level of repetition in their sequences, which allows them to be expressed in relatively large amounts in recombinant production.[140] The four small genes encoding four strands are predicted to form a tetrameric coiled coil structure.[141] The native coiled coil structure of recombinant honeybee silk protein has been reproduced using a simple biomimetic spinning system.[140] Later results showed that a single recombinant silk (AmelF 3) is sufficient to mimic the structural and mechanical property of native protein.[142] The representative structure of this single recombinant silk protein (AmelF 3) compared with silkworm silk is shown in Figure 1.9.

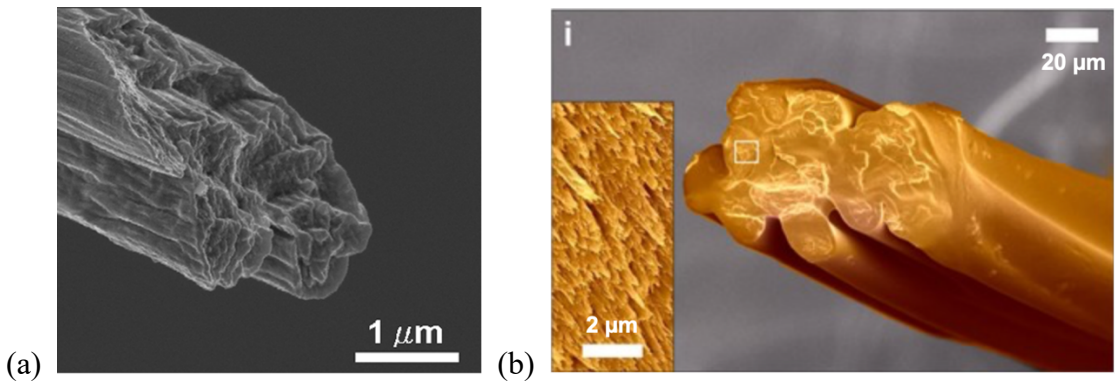


Figure 1.8– SEM images of the cross-sectional structure of (a) spider silk and (b) silkworm silk. (Reproduced from (a) Xu et al[138] and (b) Ling et al[139])

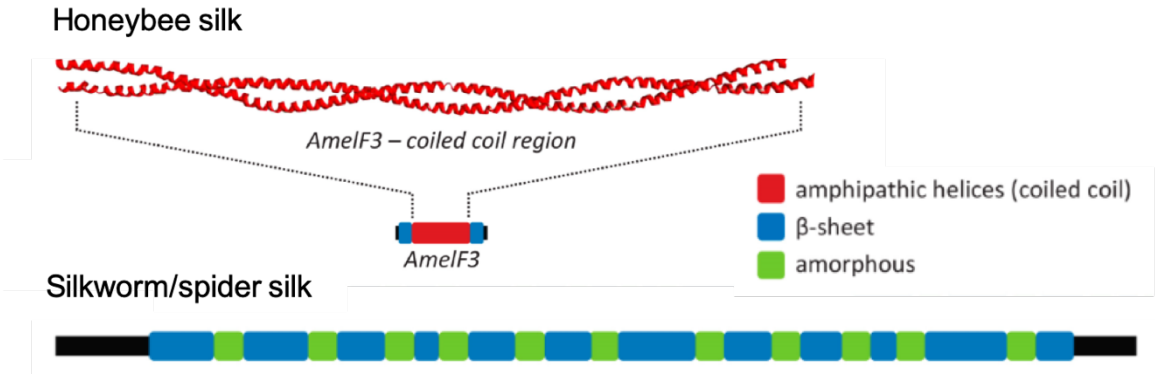


Figure 1.9– Comparison of the structure of the single honeybee silk- AmelF3, with the silkworm silk and spider silk in terms of their coiled coil and β sheet component. (Reproduced from Rapson et al)[143]

1.2.2 Application of Fibrous Materials

Collagen has been widely utilised in areas such as tissue engineering, drug delivery and biosensing. It provides potential advantages such as mechanical strength and biocompatibility upon fabrication in the form of hydrogels, fibres or sponges.[144] The mimetic of extracellular matrix (ECM) using collagen has been successful in intelligent tissue engineering applications.[145] More recently, collagen has been utilised in sensing applications based on their chemical and physical properties such as swelling, contraction on heating, high affinity with water, and its piezoelectric coefficient (8pC/N).[146–148] Researchers have reported the use of collagen as capacitor for humidity sensing.[149,150] Vivekananthan et al further utilised the

collagen as the piezoelectric nanogenerator, which has the dual functionality of energy harvesting and humidity sensing.[146] In addition, collagen-based electroconductive hydrogel has been achieved through the modification with conductive material such as carbon nanotubes and polypyrrole, which allows collagen to be in the area of electronic sensors. The immobilisation of glucose oxidase in the polypyrrole modified collagen matrix has been achieved. This self-supportive hydrogel system has been successfully used as a glucose sensor.[151] Wang et al also reported the first collagen aggregates as flexible sensors upon fusion with carbon nanotubes.[152]

Silk materials possess some superior property than collagen such as the higher elasticity. The successful recombinant production of silk protein and the various fabrication techniques allow them to be used in a larger variety of applications. Unlike the widely accepted biodegradable property of collagen, the utilisation of silk has been addressed in different areas, mostly for their *in vitro* applications. First of all, silk materials can be engineered to various formats.[153] Many researchers have reported the synthesis of silk nanoparticles for drug delivery.[154,155] The silk fibroins are regenerated to form a spherical and spindled shape. An enhanced cellular uptake and Doxorubicin delivery were observed with the reversal silk-sericin based nanoparticles.[156] Furthermore, the sensing applications for silk materials are more widely explored due to their multi-layered structures. The most common sensors developed using silk fibres include pressure, temperature, and humidity, according to the optical and mechanical properties of the silk fibres. Humidity sensing with spider silk fibres with different detection mechanisms was achieved. Liu et al utilise the diameter change of the silk fibre upon humidity change, which will change the interference-cavity change, and consequently resulting in shifts either in an interference spectrum,[157] or the resonant dips due to the refractive index change.[158] Kim et al also reported the incorporation of spider silk fibres into metamaterials. The volume change of the fibre from humidity was then detected by a change in the terahertz (THz) resonance frequencies.[159] The inclusion of gold nanostructures with nature silk materials produced a metal-insulator-metal resonator, by measuring the swelling of the insulating silk through plasmonic resonance shifts. A highly sensitive glucose sensor is realised by this method, illustrating the application of silk materials in optical biosensing area.[160]

Apart from the intrinsic mechanical properties as mentioned above, recent researchers attempt to insert novel functionalities into silk fibrous materials. The fusion of silk fibres with conductive materials could also achieve electronic sensing systems. Li et al reported a graphene engineered spider silk fibres as humidity and motion sensitivity, based on the super-contractive property of the silk materials and the conductivity change of attached/detached graphene.[161] The development of carbon fibre fused silk material, for the first time, fulfilled a temperature-pressure dual mode E-skin, which can distinguish exhaling, finger printing and spatial distribution of temperature and pressure. This achievement promised future application of the silk fibre in human machine interfaces and soft electronics.[162] Using the photolithographically patterning method with conductive ink consisting of the silk material and a conducting polymer, a biomarker immobilised biosensor was developed, which is suitable in monitoring wound healing and early disease diagnosis.[163] In addition, Singhn et al has shown that spider silk can be synthesized and assembled with gold nanoparticles to form a gold nanoparticle–spider-silk bioconjugate material which enables the study of the electrical properties of the nano-bioconjugate. By exposing the gold nanoparticle–spider silk bioconjugate to vapours of methanol and chloroform will lead to changes in electrical transport through the nanoparticles. This can be combined with the well-known contraction/expansion behaviour of the fibres in solvents of varying polarity to develop a vapor sensor.[164] In another work, even though the conductivity was not achieved unless a carbonisation process was carried out, the feeding of the silkworm with either single-walled carbon nanotube or graphene enhanced the mechanical properties of the secreted silk.[165] Furthermore, surface modifications of silk fibres by physical treatments such as UV and plasma, or chemical agents such as acid anhydride and organic dyes, allow researchers to regulate cell response and improve cell adhesion to silk substrates.[166]

1.2.3 Recombinant Silk

Similar to the recombinant protein peptides, synthetic biology also provides an important pathway for the regeneration of artificial silk materials to be used in a variety of applications, with benefits of added protein functions while maintaining the mechanical superiorities. Unlike the limited application of recombinant collagen, fusion of silk proteins (such as spider silk, silkworm silk, honeybee silk) with functional proteins has been utilised in many areas. For

example, research has illustrated functional peptides such as the antimicrobial peptide, the peptide with cellulose binding properties, and the radio protective peptide, can be fused with silks by a fermentation-based cost-structure technology to introduce functional attributes in the protein polymer thereby opening a new possibility for their use in biodefense, industrial biotechnology, and personal care, respectively.[167]

The utilisation of recombinant silk material in wound healing and tissue regeneration has been a hot topic in the 21st century. Bini et al. included RGD (Arg-Gly-Asp) cell-binding domains into the recombinant spider silk. The results show that when RGD encoded spider silk was fed to tissue culture, they could support enhanced differentiation of human bone marrow-derived mesenchymal stem cells (hMSCs) to osteogenic outcomes. This suggests that they can be used as a potential material for stem cell tissue engineering.[168] In a similar study, the spider silk was fused with the silica-binding peptide (R5), which reveals to mechanism of how the silk-silica biomaterials stimulate osteogenesis. This study provides suitable system for tissue regeneration and inspires more efficient biomaterial designs.[169] The fusion of spider silk with surface binding proteins is not unique. Horak et al demonstrated an antibody-binding domain fused spider silk protein being used in biomarker applications.[170] Recently, recombinant silk material is widely studied as a biocompatible material for wound healing. The fusion of silkworm silk with human fibroblast growth factor significantly promoted NIH/3T3 cell proliferation with no cytotoxicity.[171] The addition of the transforming growth factor in this fusion system further reduces lipopolysaccharides-induced inflammation.[172]

Kaplan and co-workers studied the recombinant silk materials in a variety of applications. First of all, they investigated the self-assembly and physical properties of recombinant silk proteins with controlled hydroxyapatite (HA) formation to make biomaterial composites by fusing with the carboxyl terminal domain of the dentin matrix protein. The protein was mineralized using simulated body fluids (SBF) and induced the formation of calcium-deficient carbonated HA which indicates it can be a potential solution in bone tissue engineering. Besides, they have also reported that recombinant silk proteins can be used for gene delivery by tailoring to contain additional features via genetic engineering.[173] During their research, they found that novel silk-based copolymer block can be realised by combining poly(L-lysine) domains to interact with plasmid DNA (pDNA) and RGD, to enhance cell-binding and transfection efficiency. The

resulting material proves silk proteins with additional functional features can be used as a new family of highly tailored gene delivery systems.[174]

Furthermore, recombinant silk may be also used in sensing applications. Brenner et al reported an intracellular tension sensor made from the spider silk peptide sequence fused in between a FRET (Fluorescent Resonance Energy Transfer) pair. The resultant spider silk peptide behaves as linear springs for the FRET pair to switch between the fluorescence on and off. This method achieved the sensing of tension strength by recombinant silk in living cells.[175] The successful production of xylanase fused spider silk with significant activity and stability provides a proof-of-concept for recombinant silk being used in a variety of enzymatic reaction systems and biosensors.[176]

1.3 Protein Assembly at the Air-Water Interface (AWI) for Fibrous Materials

1.3.1 Interfacial Assembled Protein Products

While nature creates the fibrous structures through the self-assembly of different proteins (short peptide sequences, denatured globular proteins, and scleroproteins), many techniques have been developed to direct the assembly processes, including electrospinning and a variety of printing techniques. These techniques are successfully applied to structural proteins for the production of fibres in the submicron range for a variety of applications such as wearable sensing.[177,178] Attempts on producing polymer-based functional protein fibres using these techniques have been carried out with decreased activities.[179,180] In these scenario, interfacial assembly provides an alternative pathway for immobilising proteins and making various protein scaffolds such as protein films[181–185] and protein fibres[183,184,186].

Protein molecules are amphiphilic due to their complex compositions and structures. First of all, a protein molecule consists of both hydrophilic and hydrophobic side chains. Secondly, they may be folded with either the exposure of the hydrophilic side chains or the exposure of the hydrophobic side chains depending on their functionality. The intermolecular forces such as H-bonds, hydrophobic interactions, disulphide bonds further stabilise these amphiphilic structures. A globular protein, for example, fold the hydrophobic part inside to increase its solubility in physical environment; whereas the scleroproteins expose their hydrophobic chains as they assemble in nature, and form the supportive structure. This amphiphilic property allows the interfacial adsorption of protein molecules.

One of the most common interfaces is the air-water interface (AWI). The air-water interface is a ubiquitous chemical, physical and biological entity. It is believed that water next to hydrophilic surfaces exhibit different characteristics from the water in bulk. A variety of interesting phenomenon has been found to occur at the air-water interface such as the surface tension dependent adsorption of surfactants.[187] Proteins could rearrange themselves at the interface with the hydrophilic parts remaining in the water, and the hydrophobic parts directing

towards the air, forming either uniform monolayer[183,184]/multilayer[181,182] films of protein molecules, or protein films with isolated islands (saturated films)[185].

While protein structures formed at the AWI are mainly reported for the study of protein adsorption mechanism, proteins may be integrated in other organic films to develop either electronically active materials or photosensitive optical devices. For instance, it has been shown that a stable monolayer of poly(3-hexyl thiophene) (P3HT)/stearic acid (SA), which is formed onto the air–water interface of the Langmuir trough, can be transferred successfully onto indium-tin-oxide (ITO) glass plates. Glucose oxidase (GOX), which is later immobilized on these P3HT/SA films, remains active on the film which means it can be used to make a biosensor for detecting glucose concentrations of 100–500 mg/dL.[188] In addition, Sharma et al. has found that the mixture of P3HT/SA containing galactose oxidase (GaO) may assemble at the air–water interface, forming a stable monolayer without later modifications. Similar to the simple monolayers of P3HT/SA, they can also be transferred on to the ITO-coated glass plates. These GaO immobilized P3HT/SA/LB film can be used as working electrode and Pt as reference electrode for the estimation of galactose with detection limit and saturation value were 1 and 4 g/dl galactose, respectively.[189] A further study reported an organic fibrous material generated from layer-by-layer rolling of PDMS/PSPI thin film on AWI. The fibre with a PDMS core shows its ability being used as colorimetric strain sensors.[190]

The direct formation of protein scaffolds on the AWI are limited. However, Huang et al[191] and Tsai et.al[192] have successfully developed new platforms of recombinant ultrabithorax (fused with functional proteins) to create a self-supportive, functionalized protein fibrous materials from the AWI. By using this platform, the fibrous materials are formed without damaging the structure and function of the appended proteins. In addition, this method has been tested on functional proteins with various sizes, surface charges and structures, which allows more functionalities to be added and hence extends the application areas. The research in this thesis shares many similarities with this idea but quantified the functionality with detailed study of the assembly mechanism.

1.3.2 Protein Adsorption Kinetics at the AWI

Even though protein assembly at the AWI successfully creates these interesting products as mentioned in previous section, the adsorption kinetics is complicated and attracts much research attention. The protein adsorption at AWI consists of several activities (Figure 1.10), not necessarily sequential.[193] Different spectroscopic techniques have been utilised to study the layer thickness, surface layer concentration, local mobility and folding and unfolding details of the adsorbed protein.

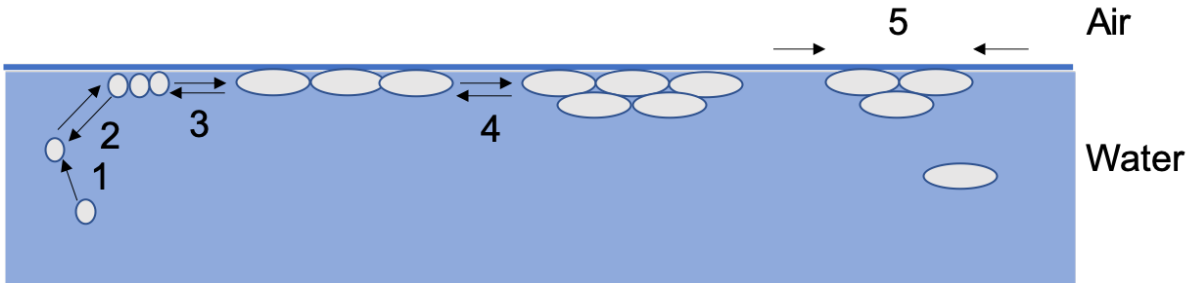


Figure 1.10– Schematic representations of protein adsorption steps at the AWI. (1) Bulk diffusion; (2) adsorption and desorption; (3) conformational change; (4) lateral diffusion and rearrangements; (5) response to applied shear stress. (Reproduced from Jongh et al)[193]

One of the most important parameters to study molecular interfacial behaviour is the surface pressure (Π). Although interpretations of the Π-t plot (Figure 1.11 (a)) were attempted by other researchers, the conclusions have been controversial in terms of the reversibility of the protein molecules adsorbed at the AWI, the interfacial energy barrier for adsorption and desorption, and the process of protein configurational changes.[194] In most explanations, a log (dΠ/dt) versus Π plot was used to visually aid the explanation of protein interfacial assembly kinetics. The plot comprises of a line with two linear regions (Figure 1.11 (b)). In earlier suggestions, researchers explained the initial linear region with a smaller slope as the penetration of molecules into the surface, whereas the absolute value of the slope increases above a transition point, corresponding to purely molecular configurational changes. The negative slopes in both regions are indicating a steady decrease in logarithm of the speed of protein adsorption to the surface, as the surface gradually becomes crowded by the protein molecules. The transition reveals a faster decrease in the logarithm of adsorption speed due to a conformational

change.[194–196] Lately, it was suggested that after the surface pressure increased due to adsorption, molecules already adsorbed on the surface adjust their configurations to attain new equilibrium for further adsorption. Therefore, the slope alteration in the second region after the transition point was attributed by both protein conformational changes, and protein further adsorption at a slower rate.[194]

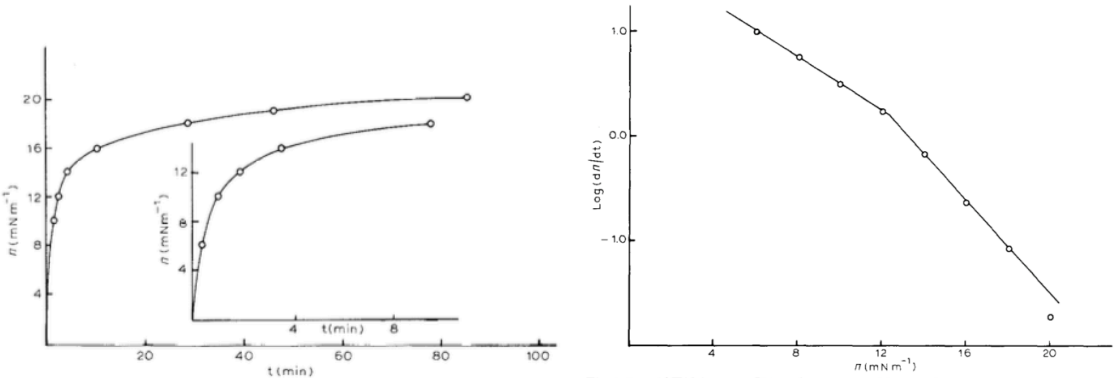


Figure 1.11– Adsorption kinetics of β-lactoglobulin at an air-water interface. (a) Plot of surface pressure (Π) over time; (b) plot of adsorption speed logarithm (log(dΠ/dt)) over surface pressure (Π). (Reproduced from MacRitchie)[194]

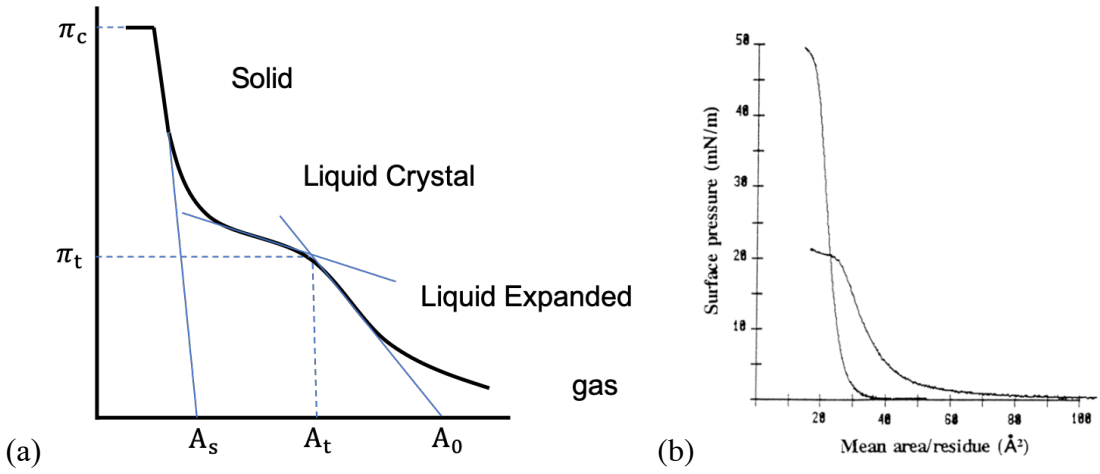


Figure 1.12– (a) Compression isotherm of a monolayer. LE: Liquid expanded state; LC: Liquid condensed state; S: solid state; πc: collapse pressure; πt: transition pressure from LE phase to LC phase; At: mean area at πt; A0: limiting area; As: area in solid state. (Reproduced from Maget-Dana R)[197] (b) Compression isotherm comparing a monolayer of β-sheet peptides (left) and a monolayer of α-helical peptides (right). (Reproduced from Maget-Dana R et al)[198]

The interpretation of a compression isotherm of proteins at the air-water interface provides more information on protein molecular footprint, protein structural change and rearrangement, and the film status upon surface compression. The compression can be controlled to result in either monolayer, multilayer films or uneven films with isolated molecular islands. Figure 1.12 (a) shows a typical compression isotherm (π-A) of monolayer film.[197] It denotes the surface pressure change over the mean molecular area, which is calculated from the total surface area and the number of molecules on the surface. A full compression isotherm starts with gaseous phase of molecules, at which status molecules do not interact with each other and gives no surface pressure change. The first surface pressure increase starts when molecules on the interface come to a liquid expanded (LE) status. The x-intercept of the linear increase is called limiting area, which gives information of this first phase transition and the compressibility of the surface. This LE state will be further transited to a liquid condensed (LC) phase starting from a relatively lower slope, as the surface is gradually compressed. The LC state begins at transition pressure π_t and the corresponding area A_t. Finally, the solid state with the lowest compressibility started and the π-A plot extrapolates to 0 at an area A_0. After this, the film

collapses at the pressure π_c. However, the surface does not usually start with a gaseous phase, and sometimes does not go through every step of transition. For example, a study comparing the compression isotherm of a monolayer of α-helix and a monolayer of β-sheet peptides was carried out, showing a higher compressibility and less stable monolayer of helical peptides, which collapse at a lower surface pressure. In both cases, the transition from LE to LC was negligible (Figure 1.12 (b)).[198] A compression of α-helix below its molecular cross-sectional area resulted a transition to β-sheet. [198,199] The assembly of protein is much more complicated, and it was also reported that the unit area occupied by β-lactoglobulin in a solid-state film at the air-water interface is ~2.1 nm^2, which is much smaller than the area of the whole molecule. In addition, it has been calculated that there is a large activation energy barrier for proteins to desorb from the surface unless the area of a molecule is reduced to a certain value, which is 1.0 nm^2 in the case of β-lactoglobulin.[200]

1.3.3 Protein Structural Change at the AWI

While protein adsorption kinetics at the AWI was well-established, the consequent loss of protein biological activity and the protein unfolding mechanisms are widely acknowledged, yet still not completely understood, due to the complexity and variety of protein structures. Intensive research has been focused on the study of peptide interfacial behaviour as they are building blocks of protein molecules, are relatively small, have much simpler and controllable structures, and thus could potentially provide information on protein behaviours. Among native and synthetic peptide sequences, amyloids are popular as the soluble α-helix rich peptides tend to form β-sheet oligomers that subsequently rearrange into fibril structures. Such structure transition property is ideal for the study of peptide arrangement and adsorption kinetics at the AWI, in comparison with their fibril-forming behaviour in bulk solution. M. Hoernke et al proposed that the amyloid peptides initially adopt an α-helical structure when they adsorbed at the interface. An increase in the local concentration at the interface triggers β-sheet formation.[201]

Mitaben et al[202] first compared the structural change of three globular proteins: lysozyme, bovine serum albumin (BSA), and β-lactoglobulin, through surface pressure measurements and

FTIR spectroscopy. They reported a slow adsorption of lysozyme with significant structural change and high content of anti-parallel β-sheet; whereas the BSA and β-lactoglobulin assembles much faster with little change in structures. They draw the conclusion that the protein experience smaller conformational change with faster assembly. Yuhko et al,[203] however, discovered a different phenomenon with X-ray reflectivity measurement. They reported a larger structural deformation of BSA comparing to lysozyme, even though the assembly of lysozyme remained the slowest among the three types of proteins. The difference might result from a higher BSA concentration used by Yuhko[203] (1 mg/ml) comparing to Mitaben[202] (0.5 mg/ml) due to the limitation of the FTIR spectroscopy. The conclusion was extended by adding the kinetics of protein unfolding into consideration, which is explained by the hydrophobic patches of the material.[203] Furthermore, researchers reported stable elastic films with fibrous textures formed by silk proteins at the AWI.[204] These fibrous silk structures on the interface experienced a structural change, which is dependent on the bulk concentration.

1.4 Aims and Objectives

This project aims at producing unique fibrous materials displaying protein functionality for sensing applications. This can be divided into two parts:

1. Molecular fusion of the protein to produce functional building blocks for assembly.
2. The design and fabrication of a protein fibre as the framework for production of functional fibrous material with potential sensing applications.

The approach taken was to use two functional building block partners with different folded structures and functions:

- Monomeric red fluorescent protein (mcRFP)
- Luciferase (Luc)

together with the assembly directing polypeptide tag:

- Polyglutamine (Q77)

to investigate the fusion effect on the functionality to develop the building blocks. Six recombinant constructs are studied with either pure fluorescence (mcRFP/mcRFP-Q77), pure bioluminescence (Luc/Luc-Q77), or dual functionalities (Luc-mcRFP/Luc-mcRFP-Q77) with potential bioluminescent resonance energy transfer (BRET) activity.

The focus is to understand the effect of fusion on the apparent functionality changes, and thus the structural insight in these changes. Self-assembly and interfacial-directed assembly is compared for the six constructs. Protein secondary structural changes are monitored during the assembly processes by CD and IR and interpretation of changes in the fluorescent properties of the constituent proteins. Finally, the functionality of the assembled product is determined by fluorescence and bioluminescence quantification. The unique mechanical properties in comparison with the fibrous structures will also be considered. Finally, the functional fibrous protein material is examined in use of ATP sensing, to achieve a proof of concept functional fibrous protein biosensor.

Chapter 2 Materials and Methods

2.1 Materials

For protein expression and purification experiments, isopropyl β-D-1-thiogalactopyranoside (IPTG), TritonX-100, phosphate buffered saline (PBS), sodium chloride (NaCl), sodium phosphate monobasic and dibasic ($NaHPO_4$, Na_2HPO_4), lysozyme (from chicken egg white), imidazole, trisaminomethane hydrochloride (Tris-HCl), ethylenediaminetetraacetic acid (EDTA), 1,4-dithiothreitol (DTT), EDTA-free protease inhibitor cocktail and benzonase nuclease®, ammonium persulfate (APS), N, N, N, N- Tetramethylethylenediamine (TEMED), β-mercaptoethanol, and sodium dodecyl sulfate (SDS) were purchased from Sigma Aldrich. Quick Coomassie stain is sourced from Generon; Sephadex G-25 in PD-10 desalting columns from GE healthcare; 0.22 μm pore size PVDF syringe filters from New England Biolabs. Pierce™ unstained protein MW marker, HisPurTM Ni-NTA resin were purchased from ThermoFisher Scientific, and magnesium sulphate from Fisons. Vivaspin® Turbo 15 (30 Dalton) was obtained from Satorious. Luria Bertani Medium was made using 10g Bacto-tryptone, 5g Bacto-yeast extract and 10g NaCl in 1 litre distilled water. Kanamycin and ampicillin powder were purchased from Melford Laboratories Ltd. A stock concentration of 50 mg/mL was prepared in milli-Q water for each of them.

D-luciferin, adenosine triphosphate (ATP), thioflavin-T (ThT), urea, ethanol, acetone, and isopropanol were purchased from Sigma Aldrich. Pierce™ BCA protein assay kit were obtained from ThermoFisher Scientific. Unless otherwise stated, all water used in this thesis is UHP (ultrahigh pure) water sterilised from milli-Q water with a resistance of 18.2 MΩ.

2.2 Isolation of Recombinant Proteins

2.2.1 DNA Constructs of Fusion Protein Models

Different combination of the two functional building block partners with the assembly directing polypeptide tag results in 6 fusion constructs of interest, which are denoted as: mcRFP, mcRFP-Q77, Luc, Luc-Q77, Luc-mcRFP, and Luc-mcRFP-Q77. All constructs with 6x histidine tag (Figure 2.1) were previously developed in Hall's laboratory (Cambridge Analytical Biotechnology Group, University of Cambridge, UK). mcRFP in the pET-24(a)+ plasmid (Novagen) was designed previously in the laboratory by modifying a stock of the mRFP1 developed by Campbell et al.[93] It has a C terminal His-tag for Ni^{2+} purification, and one serine to cysteine mutation at the position of 131 for increased spectral intensity. The polyglutamine sequence- Q79 in a commercial plasmid pGEX (GE Healthcare was a kind gift of Prof. Akira Kakizuka (Laboratory of Functional Biology, Kyoto University, Japan). Q79 indicates the number of polyglutamine residues in tandem. Previous sequencing of the gifted plasmids revealed glutamine loss where pGEX-Q79 contained 77 glutamine residues in tandem, apart from one lysine at position 4 within the Q sequence. These polyglutamine sequences were amplified with the help of sequence specific primers using PCR and subsequently cloned into pET-24(a)+ (Novagen) between EcoR1 and Xho1 restriction sites by Dr. Muhammad Safwan Akram. mcRFP-Q77 was also previously developed by cloning the mcRFP sequence to the N terminal of the Q77 plasmid.[205]

Similarly, firefly luciferase with 5-point mutations from the wild type in pET-16b was produced by Dr. G. H. Erica Law [101] and stored in house. The gene for firefly luciferase was amplified and cloned to the previously produced Q77, mcRFP, or mcRFP-Q77 in pET-24(a)+ plasmids by Dr. J. U. Rehman.[206] All constructs have N-terminal hexa-His tags. The full DNA constructs are shown in Appendix A.1.

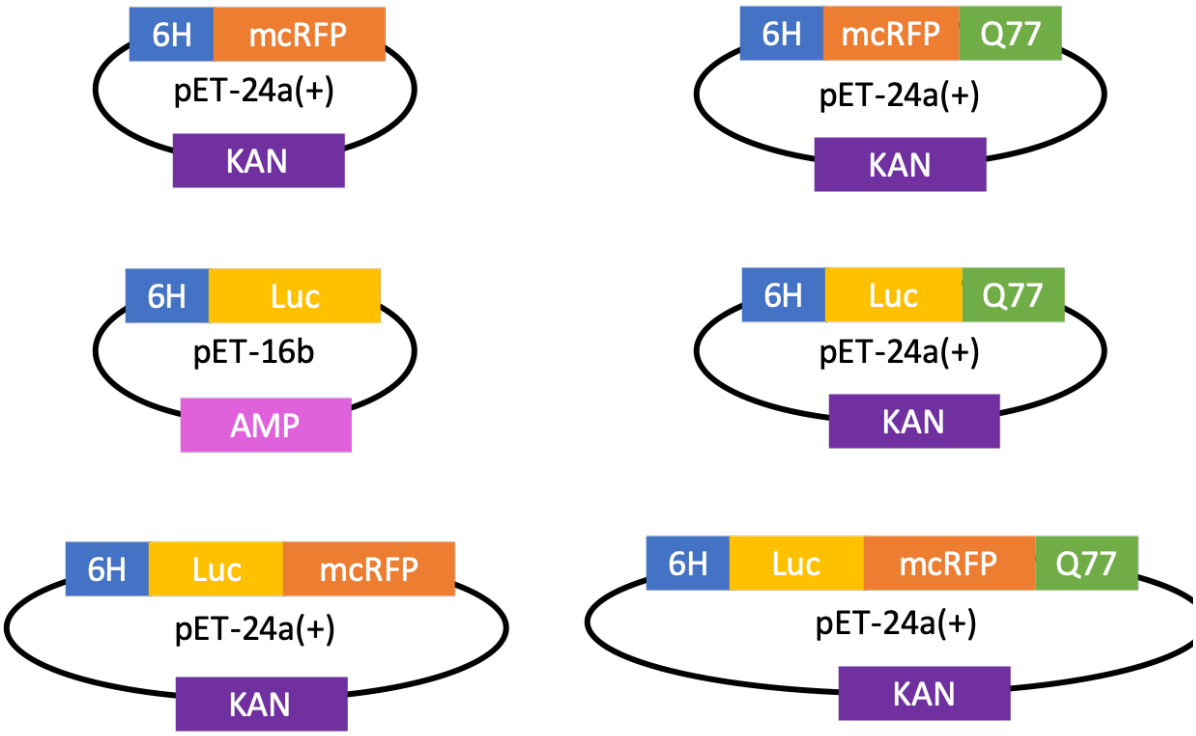


Figure 2.1– Plasmid compilation structures of each constructs with plasmid resistant site.

2.2.2 Protein Expression and Cell Lysis

2.2.2.1 mcRFP and mcRFP-Q77

The competent *E. coli* BL21 (DE3) cells with mcRFP and mcRFP-Q77 plasmids were stored as the glycerol stock, made from 500 µL of overnight cell culture and 500 µL of 60% glycerol in LB broth. Cells picked from the glycerol stock were grown in LB culture (10 mL) supplemented with kanamycin (100 µg/mL) for 12-16 hours as the overnight cell culture. The culture was diluted 1:100 in 400 mL LB media with kanamycin and further incubated until an $O.D_{600}$ of 0.6-0.8 was reached. Protein expression was induced with 1mM IPTG. All cultures were grown at 37 °C with shaking at 225 rpm.

Cells were harvested by centrifugation at 4000 rpm at 4 °C for 20 minutes and resuspended in lysis buffer (0.5 M NaCl, 20 mM Tris-HCl, 5 mM imidazole, 100 units of lysozyme) by votex. The suspension was left on the bench at room temperature for 40 minutes. Interactions between cells and the extracts were broken with probe sonication for 30 cycles. Each cycle consisted of

alternating periods of 10 seconds on and 5 seconds off. The solution was centrifuged at 13,500 rpm for 45 minutes at 4 ºC. Cell debris was collected at the bottom while the supernatant containing the proteins was isolated for downstream processing.

2.2.2.2 Luc, Luc-Q77, Luc-mcRFP and Luc-mcRFP-Q77

Luciferase, Luciferase-mcRFP, luciferase-Q77 and Luciferase-mcRFP-Q77 plasmids were transformed into competent E. coli Rosetta 2 cells (a derivative of BL21 cells designed for enhanced protein expression for eukaryotic proteins by supplying tRNAs for 7 rare codons). Cells were grown in LB culture (10 mL) supplemented with kanamycin (100 µg/mL, for LR, LQ, and LRQ constructs) or ampicillin (100 µg/mL, for luciferase construct) for 12-16 hours. The starting culture was diluted 1:100 in 200 mL LB media with the appropriate antibiotics and further incubated until an $O.D_{600}$ of 0.6-0.8 was reached. Protein expression was induced with 1mM IPTG. All cultures were grown at 37 °C with shaking at 225 rpm.

Cells were re-suspended in buffer A (300 mM NaCl, 2.7 mM KCl, 10 mM sodium hydrogen phosphate, 1.76 mM potassium dihydrogen phosphate, 10 mM 2-Mercaptoethanol, 0.1% protease inhibitor cocktail and 20% glycerol, pH 8.0) supplemented with 2% TritonX-100 and 20 mM imidazole. The pellet was dissolved by shaking with a spoon. Resuspended cells were treated with 3 cycles of freeze and thaw at -80 ºC. Benzonase nuclease and lysozyme was finally added to the cell extracts. The solution was centrifuged at 13,500 rpm for 45 minutes at 4 ºC. Cell debris was collected at the bottom while the supernatant containing the proteins was isolated for downstream processing.

2.2.3 Protein Purification and Identification

Protein purification was performed using Nickel His-Bind Resin (Novagen). The purification columns were prepared as recommended by the manufacturer. The column was first washed by 5x bed volume of equilibrium buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.8), followed by the loading of the protein supernatant. The impurities were removed by washing the column with 3x bed volume of binding buffer (buffer A with 20 mM imidazole), 5x bed volume of wash buffer (buffer A with 50 mM imidazole). The target protein

was then eluted with 5x bed volume of elution buffer (buffer A with 250 mM imidazole). The purified protein was desalted using a desalting column to exchange the elution buffer with the desired protein storage buffer [10 mM sodium phosphate buffer, pH 8 for mcRFP constructs and 1x TEM (10 mM Tris/acetate, pH 7.8, 1 mM $MgSO_4$ and 0.2 mM EDTA) containing 10% (v/v) glycerol and 2 mM DTT for luciferase constructs]. The purified proteins were confirmed using size analyses of denatured proteins on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

SDS-PAGE gel was cast by pouring resolving gel solution (1.34 mL 30% acryl-bisacryl amide stock, 1.25 mL 4X resolving gel buffer, 2.41 mL distilled water, 60 μL 10% (w/v) ammonium persulfate (APS), and 5 μL tetramethylethylenediamine (TEMED)) 6% and 8% and stacking gel solution (0.415 mL 30% acryl-bisacryl amide stock, 0.625 mL 4X stacking gel buffer, 1.46 mL distilled water, 37.5 μL 10 % (w/v) APS, and 4 μL TEMED) between two glass slides. 4X resolving gel buffer (30 g Tris-base, 144 g Glycine, 100 mL 10 % SDS stock, distilled water added to 1 L, pH 8.3) and 4X stacking buffer (30.28 g Tris base, 20 mL 10% SDS stock, distilled water added to 500 mL, pH 6.8) were prepared in advance and refrigerated. Protein samples were mixed with 4X SDS sample buffer (25 mM Tris-HCl pH 6.8, 10% sodium dodecyl sulfate (SDS), 0.004% Bromophenol blue, 20% β-mercaptoethanol, and 40% glycerol) in 4:1 ratio and incubated at 95 ºC for 5 minutes. 8 μL of protein samples were loaded into each lane of the SDS-PAGE gel along with 5 μL protein marker and the gel was run at 180 V until the front line of the dye entered the edge of the resolving gel. The gel was separated from the glass slides and incubated at 4 ºC in the box containing coomassie blue staining solution for 15 minutes. The gel was then washed with sterile UHP water and preserved in 10% methanol.

The identifications of the protein were further confirmed by MALDI mass spectrometry (Appendix A.2). The proteins were identified by peptide mass fingerprinting using trypsin digestion with the PNAC (Protein and Nucleic Acid Chemistry) facility at the Department of Biochemistry, University of Cambridge.

2.3 Activity Characterisation of Recombinant Proteins

2.3.1 Absorption and Fluorescence Spectroscopy

Fluorescence spectra were taken with the Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies). Black 96 well plates were used to carry samples with 200 μL solution in each well. For the acquisition of fluorescent spectra from different mcRFP constructs, excitation wavelength was set at 586 nm and fluorescent emission data was recorded from 596 nm to 700 nm. The excitation and emission slits were 5 nm, and the PMT (photomultiplier) voltage was set to 600 V.

The absorption spectra of proteins were obtained with a UV/Vis LAMBDA 16 spectrometer (Perkin Elmer). Transparent 96 well plates were used to carry samples with 200 μL solution in each well. For the acquisition of absorption spectra from different mcRFP constructs, the absorption was recorded from 400 nm to 700 nm. The scanning steps were set as 1 nm.

2.3.2 Quantum Yield (QY) Calculation

The quantum yield of fluorescent protein molecules produced in this thesis (mcRFP, mcRFP-Q77, Luc-mcRFP, and Luc-mcRFP-Q77) at their native state in solution was calculated by equation (2.1):

$$Q_{ref} = Q_{sample}\left(\frac{G_{ref}}{G_{sample}}\right)\left(\frac{\eta_{ref}^2}{\eta_{sample}^2}\right) \quad (2.1)$$

Where subscripts ref and sample denote for reference molecule and protein sample, respectively. Q is the quantum yield. G is the gradient of integrated emission intensity to integrated absorbance intensity at the excitation wavelength. And η is the refractive index of the solvent. The criteria for the reference was to ensure that it absorbs at the similar range of excitation wavelengths with the protein of interest. Rhodamine 101 absorbs at 565 nm, which fits to the excitation wavelength maximum of mcRFP at 586 nm. Therefore, Rhodamine 101

with QY known as 0.96 in basic ethanol at room temperature,[207] was chosen as the reference molecule for calculating QY of mcRFP molecules in each constructs.

Five reference solutions with increasing concentrations were recorded for their absorbance intensities at 564 nm, and emission maximum at 588 nm excited at 564 nm. The gradient was obtained from the trend line of the plotted integrated emission intensity against the integrated absorbance intensity (Figure 2.2). This gradient was cross calibrated by another fluorescent molecule, rhodamine B, which absorbs at lower wavelength (550 nm) and emits maximum at 570 nm. The QY of rhodamine B was reported as 0.65 in basic ethanol at room temperature.[208] Cross-referencing of rhodamine 101 and rhodamine B resulted in a gradient ratio ($\frac{G_{Rh101}}{G_{RhB}}$) of 1.48, which is within 10% variation of the QY ratio ($\frac{Q_{Rh101}}{Q_{RhB}}$=1.47) in the same solvent, validating the referencing method according to equation (2.1). The QY of protein samples were calculated similarly from the gradient of five solutions with increasing concentrations.

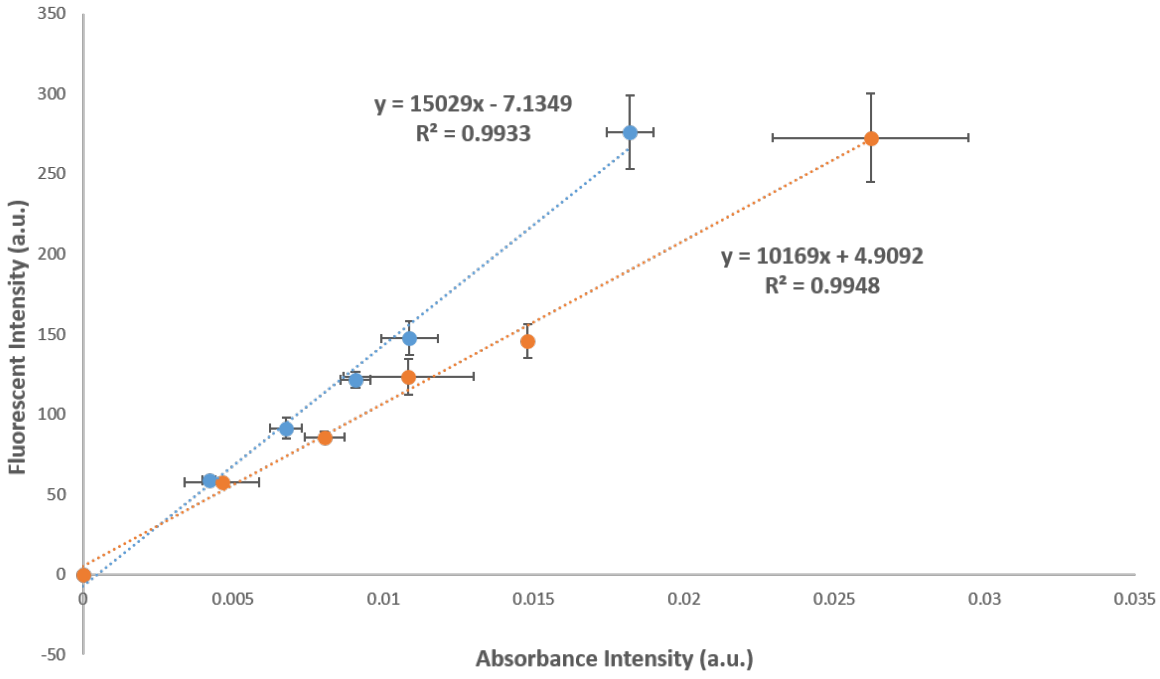


Figure 2.2– The emission intensity- absorbance intensity plot of references for QY calculation.

2.3.3 Fluorescent Lifetime

Fluorescence lifetime was measured using the mini-τ with a picosecond pulsed light emitting diode (EPLED-560, Edinburgh Instruments). The time range was set to 50 ns for a total of 1024 channels. The instrument response was tested with Sterile Ultra High Pure water with emission filter set at 1 (no filter). Protein solutions were diluted to 1 μM before each measurement. 700 μL protein solution was added to a glass cuvette for measurement. When measuring the lifetime of the protein fibres, 500 μL of buffer was added to the glass cuvette, so that the fibre is floating on the water surface within the light path. The emission filter used was 5 (filter range from 570 nm to 630 nm). Intensity was adjusted until the stop rate was stable and did not exceed 100,000 counts per second. Lifetime of samples were tested with the stop condition of 1000 counts. Data was analysed with re-convolution fit function in the F980 software.

2.3.4 Bioluminescent Spectroscopy

The specific activity of each luciferase construct was calculated by recording the bioluminescent intensity over time (using a Labsystems Luminoskan Ascent luminometer and a black 96 well plate) when 90 μL of substrate with a final concentration of 1 mM ATP and 200 μM D-LH_2 was added to 10 μL of 50 pM protein solution, or a piece of protein fibre with 10 μL storage buffer. The flash height was recorded in Relative Light Unit (RLU). PMT voltage was set at 800 V. All assays were performed in 1xTEM buffer (100 mM Tris/Acetate, pH 7.8, 2 mM $MgSO_4$ and 0.2 mM EDTA) containing 10% (v/v) glycerol and 2 mM DTT.

2.3.5 Luciferase Kinetics

Enzyme reaction involves an enzyme (E) binding to a substrate (S) to form a complex (ES), which is then converted to the original enzyme (E) and a product (P). The enzyme reaction can be described by the following equilibrium equations:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_{cat}} E + P \tag{2.2}$$

Where k_1 and k_{-1} are the rate constants for substrate binding to the enzyme and substrate unbinding, respectively; k_{cat} is the turn-over number. The turn-over number is defined as the maximum number of conversions of the substrate molecules per second that a single catalytic site will execute for a given enzyme concentration. The enzyme kinetics can be calculated by the Michaelis-Menten equations (2.3)(2.4).

$$k_{cat} = \frac{V_{max}}{[E]_{total}} \tag{2.3}$$

$$v = \frac{V_{max}[S]}{K_m + [S]} \tag{2.4}$$

Where V_{max} is the maximum rate, $[E]_{total}$ is the total enzyme concentration, v is the reaction rate, and K_m is the Michaelis-Menten constant. K_m is defined as the substrate concentration at which the reaction rate is half of the V_{max}. It can be used to measure the binding affinity of a certain substrate to a given enzyme. The ratio of k_{cat}/K_m is a measure of catalytic efficiency, which is the efficiency of an enzyme to covert a substrate to a product.[209] Generally, the enzyme-catalysed reaction is monitored by either the decrease of the substrate concentration or the increase of the product concentration, through the accumulative colorimetric method.

The K_m value of Luc protein constructs for ATP and D-LH_2 were calculated using the method described by Law et al.[101] Briefly, one substrate was kept in excess while the other was varied, and the luminescent intensity flash heights recorded. Unlike normal enzymes, whose reaction rate is indicated by the slope of the accumulative product increase. The monitoring of the luciferase catalysed reaction is through the production of light, which is non-accumulative. Therefore, the peak light emissions were taken as a measure of the initial velocity and expressed as RLU per second.[210] The maximum turn-over rate in RLU during the reaction versus the substrate concentration was plotted for each protein construct. Non-linear regression-based curve fitting, following the Michaelis-Menten model (equation (2.2)), was carried out on the resulting plot using GraphPad Prism software, yielding K_m and k_{cat} values. For these experiments, the varied concentration was chosen in the range between 0.1 and 10 times of the reported K_M value. When calculating K_m for ATP, ATP concentration was varied from 7 μM to 620 μM, while D-LH_2 was held in excess at 200 μM. For D-luciferin, D-LH_2 was varied

from 1.6 to 139 µM while ATP was held in excess at 1 mM. 10 µL of 2.5 µM protein was added to 90 µL of each substrate solution. The PMT voltages for ATP and D-LH_2 experiments were set at 400 V and 550 V, respectively, so that the maximum flash height is within the upper limit of detection.

2.3.6 Bioluminescent Resonance Energy Transfer (BRET) Measurements

BRET happens from a bioluminescent donor to an acceptor, with the distance between them of less than 10 nm, in the presence of a substrate. BRET required an overlap between the emission wavelength of the bioluminescent donor with the excitation wavelength of the receptor. A decrease in the bioluminescent intensity of the donor and an increase in the fluorescent intensity is observed due to the BRET reaction. The ratio between the fluorescent intensity and bioluminescent intensity is used to present the BRET efficiency. The BRET effects were measured by recording the emission intensity over wavelength using the Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies) with excitation light kept off, 22 seconds after the addition of 90 µL substrate with a final concentration of 1 mM ATP and 200 µM D-LH_2 to 10 µL protein. The efficiency of BRET was studied by measuring the emission intensity at 610 nm of 1 µM Luc with various concentrations of mcRFP after 22 seconds and 53 seconds of substrate addition. Total reaction volume was kept at 100 µL; and PMT voltage was set at 800 V.

2.4 Structural Determination of Recombinant Proteins

2.4.1 Fourier-Transform Infrared Spectroscopy (FT-IR)

Infrared spectra were measured using a Perkin Elmer Spectrum One FT-IR spectrometer (Perkin Elmer Inc., Waltham, USA). Corresponding buffers were used as the background for solution spectra, whereas blank carbon tape was taken as the background for film FT-IR spectra. The scanning range was set from 4000 to 600 cm^{-1} in transmittance mode with 1500 – 1700 cm^{-1} as the range of interest (amide I band for proteins).

2.4.2 Circular Dichroism (CD)

CD spectra were obtained using circular dichroism spectrometer Aviv Model 410 (Aviv biomedical, Inc., Lakewood, USA). 0.2 mg/mL protein was prepared in 10 mM sodium phosphate buffer pH 8 without any chloride ions presented. 10 mM sodium phosphate buffer was loaded into a 1 nm path-length cuvette for baseline scanning and the solution was subsequently used to clean the cuvette between sample measurements. Every protein sample was scanned 3 times from 185 nm to 260 nm with 1 nm bandwidth, 1 nm wavelength step and 3 second averaging time. Collected data were baseline corrected and smoothed using Aviv 410 software. Protein secondary structure were determined using the program CDSSTR from Dichroweb (online analysis server for protein Circular Dichroism spectra) by averaging from three different reference set: SMP 180 (reference with 28 membrane proteins), SP 175 (reference with 71 soluble proteins), and set 7 (reference with 5 denatured proteins). The percentage of each secondary structural type was calculated by fitting the experimental data with the composition of each fingerprint curve shown in Figure 2.3.

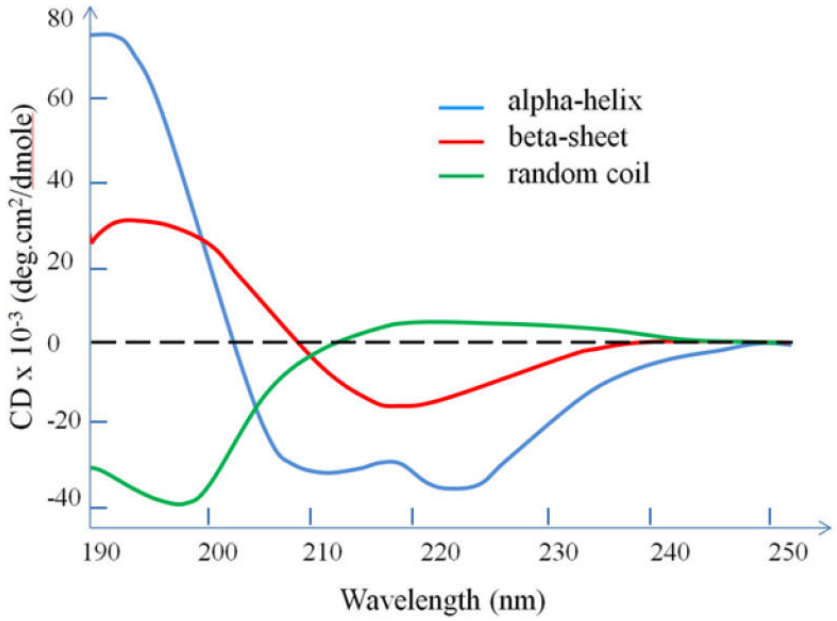


Figure 2.3– CD spectrum of each secondary structural type.[211]

2.4.3 Protein Structure Predictions

The secondary structures of the proteins were predicted with the Robetta online server, and displayed with Pymol. The structure component (α-helix and β-sheet) of each protein was calculated with a line of Python code simulated in Pymol:

ss= [i.ss for i in cmd.get_model ("n. ca").atom]

print "Helix content: %.2f%%" % (100.0*ss.count("H")/len(ss))

print "Sheet content: %.2f%%" % (100.0*ss.count("H")/len(ss))

2.5 Protein Self-assembly and Thioflavin-T Binding

Proteins were incubated at either room temperature or an evaluated temperature in the InnovaTM 4300 incubator. pH was adjusted with either 0.1 M HCl or NaOH. The formation of amyloid-like fibril structures was detected by Thioflavin-T (ThT). The structure of ThT is shown in Figure 2.4. The carbon-carbon bonds between the benzene group and the thiazole group in the ThT molecule are originally free to rotate, and consequently quenches the conjugated π-bond fluorescence of ThT with emission wavelength at 485 nm and excitation wavelength at 450 nm. The benzothiazole structure in ThT, however, could be stacked into the β-sheet structure of amyloid fibres, preventing the molecular rotation, and thus inhibits the quenching effect.[212]

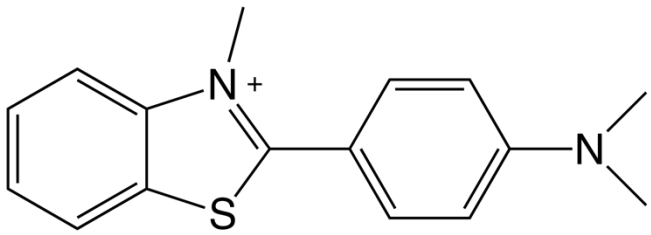


Figure 2.4– Molecular structure of ThT.

The assembled aggregates were mixed with 2.5 μM ThT in 50 mM glycine. After vortexing and incubating the mixture in dark for 5 min, the fluorescence emission was measured at 485 nm with an excitation at 450 nm using a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies). The isoelectric point (IEP) of mcRFP and mcRFP-Q77 was calculated using an online isoelectric point calculator (IPC).[213]

2.6 Microscopic and Spectroscopic Methods

2.6.1 Optical and Fluorescent Microscopy

Optical microscopic pictures were obtained by a Nikon's Optiphot-2 fluorescence microscope system (x20, x40, x63, 1.4 N.A. Plan Apochromat objective) with Digital Sight DS-2MV. For fluorescent images, the Nikon G-2A filter set and an Epi-FI illuminator were employed. The filter set was designed as the standard combination for use with green excitation wavelengths and employed a wide (50 nm) excitation pass band that covered essentially the entire green spectral region, ranging from 510 to 560 nm. The filter enabled detection of mcRFP, which has emission maxima at 607 nm.

2.6.2 Scanning Electron Microscopy

Nanoscale microscopic images were obtained with a LEO GEMINI 1530VP FEG-SEM system in the Nanoscience Centre, University of Cambridge. The machine has inlens, secondary electron and quadrant back scatter detector. The electron beam energy (represented by EHT voltage) range of the machine is 200 V to 30 kV.

Different from traditional optical microscopy, SEM uses an electron beam to interact with the samples. The electrons are accelerated under a certain potential before they reach the sample. A lower accelerating voltage will result in a secondary electron mode, where the brightness difference in the microscopic image illustrates the morphology of the surface. A hill on the surface forms a bright point, whereas a hole in the sample gives a dark point. A higher accelerating voltage will lead to a backscattered electron mode. The brightness in the image tells about either the density of the sample or different elements forming the surface. In this case, a brighter point is resulted from a higher density area or an element with larger molecular weight, and vice versa. In this work, 5 kV was used for all SEM images, which results in a secondary electron mode.

2.7 Langmuir Blodgett Films

Protein assembly was achieved at the air-water interface in a round Langmuir Blodgett trough (Nima LB2000) with a maximum volume of 1.5 L, and a surface area of 986 cm^2. The trough was first cleaned thoroughly with acetone and iso-propanol. Sterile Ultra High Pure water was added. The surface pressure was monitored by the Wihelmey plate method with a Nima surface pressure sensor. The surface pressure was monitored over one hour and then zeroed again to ensure that the cleaning solvents have evaporated completely, and the surface pressure was stable. A nozzle connected to a pump was used to remove contaminants on the surface until there was no increase in surface pressure when the trough barriers were closed. The barriers were then opened to maximum position and the surface pressure zeroed. Protein solutions were spread on the interface drop by drop with a 20 μL pipette. The amount of the protein, the type of the protein and the initial protein concentration were varied to compare the effect of each parameter. The surface pressure change was monitored for 24 hours. The surface was then compressed by closing the barrier at a speed of 100 cm/min and a plot of pressure-area isotherm was obtained.

Protein film was collected by dipping a b-shaped wire into the water (Figure 2.5 (a)). The wire was slowly lifted from the water. The protein film being tracked in the circle area of the wire were stable in air for several minutes. The film was then quickly transferred onto either glass slides for optical microscope analysis or carbon tape for SEM and FT-IR analysis.

Protein fibres were lifted by a pair of tweezers from the air-water interface. A slide of double-sided tape was stuck onto an aluminium foil covered petri dish lid. After the fibres were dried in air, one end of the fibre was stuck onto the double-sided tape and the rest of the fibre loosely stayed on the aluminium foil. This aluminium foil has electrostatic attraction of the fibre and protects the fibre from blowing away (Figure 2.5 (b)). The fibres were then easily cut-off from the tape and transferred for further application.

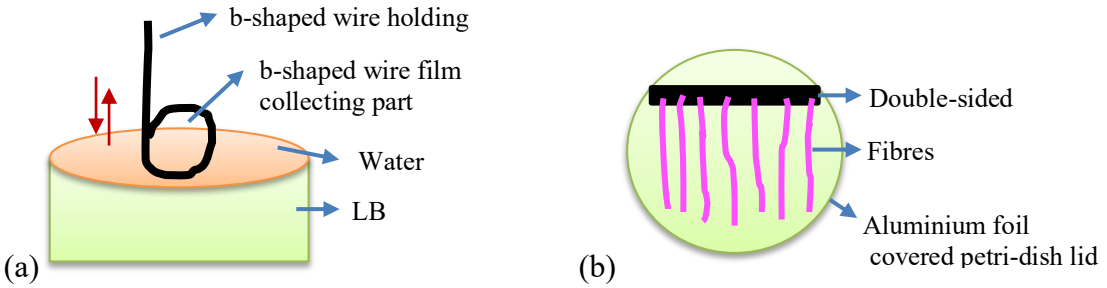


Figure 2.5– Illustration of (a) protein film collection design and (b) protein fibre collection design.

2.8 Mechanical Strength Tests

Protein fibres were collected and stuck across a 1 cm x 1 cm U-shape cardboard with superglue for mechanical strength test. The stress-strain tests were performed with the help of Prof. Eugene Terentjev at Physics of Medicine, Cavendish Laboratory. The cardboard with fibre attached was mounted on to a home-built computer controlled uniaxial stretcher (shown in Figure 2.6). The spine of the U-shape cardboard was cut so that only the fibre was connected to the stretcher's clamps. A 25 g UF1 low range isometric force sensor (LCM systems) was used to detect the force applied on the stretcher.

The system was first calibrated by attaching known weights to the force gauge and measuring the voltage response. The relationship between the voltage and force values was obtained by a linear regression fit (Figure 2.7). All strains to the fibres were applied at a crosshead speed of 0.1 mm/s until fibre breakage. Readings were recorded every 0.1 s. Force readings were converted to Stress via dividing by the cross-sectional area of the fibre. The cross-sectional area was calculated by measuring the diameter of the fibre at 10 different points across the fibre using an optical microscope. A circular cross section was assumed. Strain was calculated by subtracting the length at a time point from the initial length before stretching. These stress-strain readings for every time point were then plotted into a stress-strain curve. An example of the stress-strain curve was shown in Figure 2.8. The elastic Young's modulus was found by calculating the initial linear slope of each curve. The breaking stress and strain were the stress and strain values at the point the fibre ruptured.

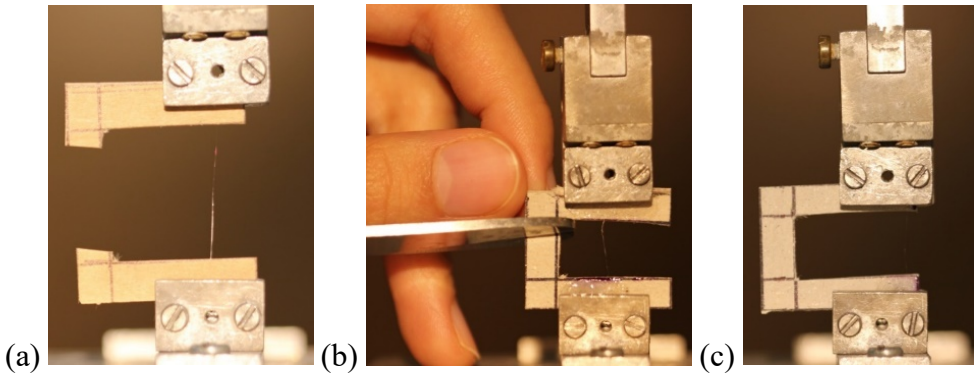


Figure 2.6– Illustration of the U-shaped cardboard attachment to the clamps of the home-build computer controlled uniaxial stretcher. (a) U-shaped cardboard loading protein fibre being clamped in the stretcher. (b) Cutting progress of the spine of cardboard. (c) Initial state before stretching and measuring.

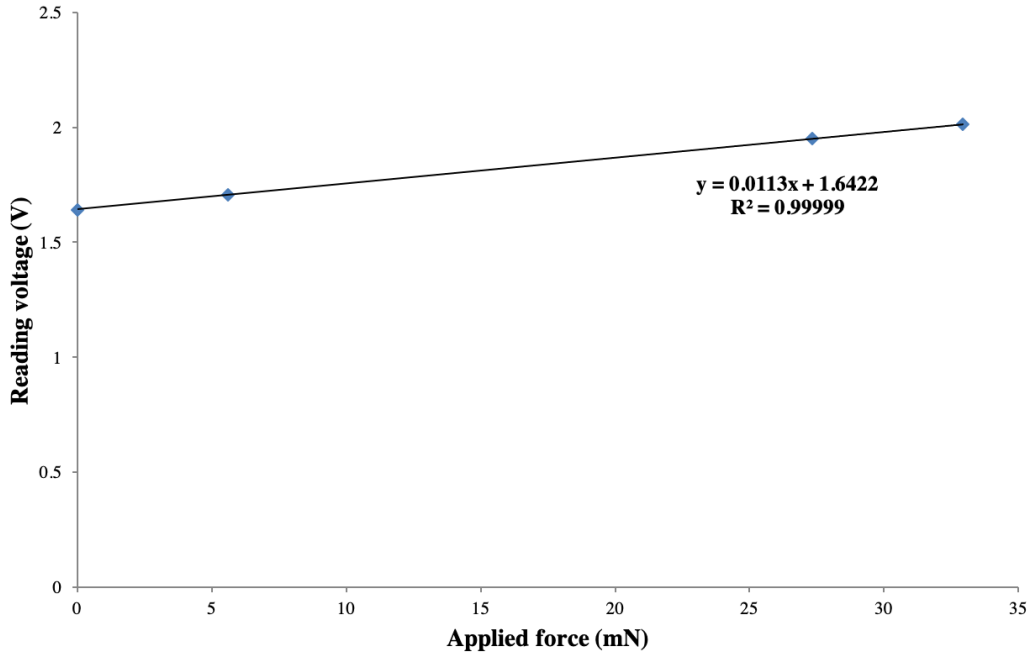


Figure 2.7– Calibration of the voltage response with known applied force.

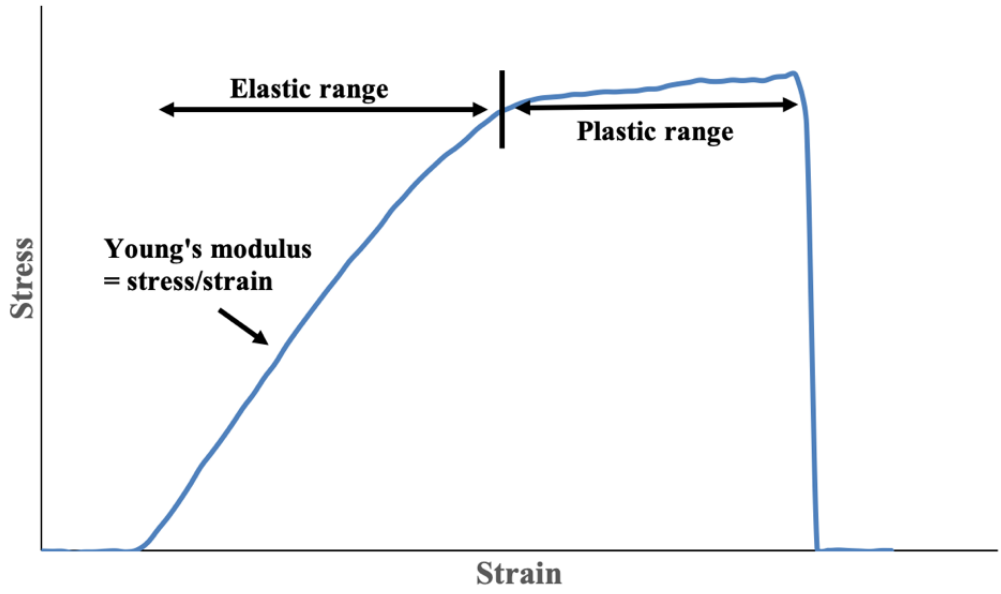


Figure 2.8– Representative plot of a typical stress-strain curve for plastic materials.

2.9 Protein Fibre Quantifications and Activity Assays

2.9.1 Protein Fibre Quantifications

2.9.1.1 Modified Bicinchoninic Acid (BCA) Assay for mcRFP Protein Fibres

The first method to evaluate the total protein amount in fibres was using the Thermo Scientific™ Pierce™ BCA protein assay. This method is chosen as it produces a soluble final product whose amount is proportional to the total amount of protein reacted, rather than forming a complex with the protein such as the widely-used Bradford assay. Peptides containing three or more amino acid residues form a coloured chelate complex with cupric ions in an alkaline environment containing sodium potassium tartrate, and reduces the cupric ion (Cu^{2+}) to cuprous ion (Cu^{+}). This step is known as the biuret reaction. Bicinchonic acid (BCA) then reacts with Cu^{+} to form an intense purple-coloured product. This BCA/copper complex is water-soluble and could be detected by measuring the absorbance at 562 nm. To evaluate this method, a standard curve was taken by varying the concentration of bovine serum albumin (BSA), shown in Figure 2.9. mcRFP and mcRFP-Q77 fibres were treated in the same way as BSA for the initial estimation of total protein. However, in order to make an accurate estimate of total protein in the fibre, protein fibres were first treated with 80 μL of 8 M urea, which denatured the protein structure but dissolved the protein fibres. Protein solutions were first prepared by 5 times of serial dilution to a final volume of 40 μL each. Another 40 μL of 16 M urea was added to each protein diluent. These urea-treated protein solutions were used as the standard samples for the corresponding protein fibres. Three replicates of 25 μL protein standards (BSA or urea denatured protein solutions) or 25 μL protein fibre samples were mixed with 200 μL working reagent in a 96 well plate. The mixture was then shaken at 860 rpm for 30 seconds using a plate shaker before being incubated at 37 °C for 30 minutes and 50 °C for 60 minutes for BSA and the urea denatured proteins (solution or fibre), respectively. The incubation temperature and time were increased for the urea denatured proteins to increase the sensitivity for these standard samples. A standard curve was performed for each batch of fibre test to prevent errors from batch to batch.

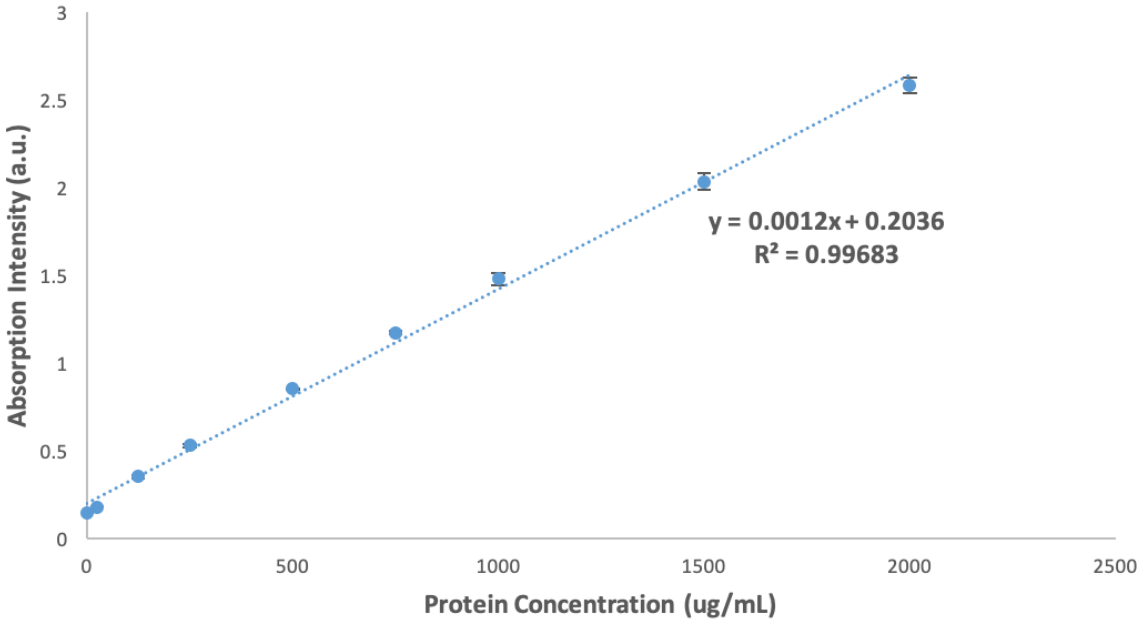


Figure 2.9– BCA standard assay of bovine serum albumin (BSA).

2.9.1.2 Volume Calculation for Luciferase Protein Fibres

Another method to quantify the total protein in fibres was using the volume of the fibre. This method, however, assumes that the packing density in each specific type of the fibre remained the same. This is validated since the fibres of each type were produced from either the same film, or films formed under the same conditions at the air-water interface. The optical microscopic pictures of each part of the fibre were obtained using the Nikon microscope with the lens set to 4x. The pictures were joined together in Microsoft Powerpoint and saved as full fibre pictures with scale. The fibre was assumed to be cylindrical and the diameter was measured by taking pictures at 10 different positions over the fibre, each for roughly 1 cm length, using the Nikon microscope with the lens set to 20x. The length and the diameter of the fibre were then calculated with imageJ. And the total volume of the fibre was calculated as:

$$V_f = L_f \pi \left(\frac{d_f}{2}\right)^2 \tag{2.5}$$

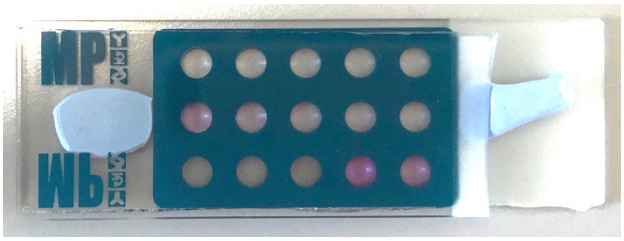
where V_f is the total volume of the fibre in Å^3, L_f is the total length of the fibre, and d_f is the diameter of the fibre in Å.

2.9.2 Fibre Activity Determinations

2.9.2.1 BiognostiX Reader for mcRFP Protein Fibres

In order to quantify the active red fluorescent molecules in mcRFP fibres, an imaging method is developed with a home-build BiognostiX Reader, which excites fluorescent molecules on the holding slide and scans the fluorescent image with a magnification of 0.38. The fibres were imaged on a glass holder (cut into the same size as the holding slide) with a cover slide. The fluorescent intensity of mcRFP solution under the BiognostiX Reader was used as a calibration to calculate the amount of protein in the fibre. 20 µL protein solution was dropped in a well on a multitest glass slide, the two blank edges of the slide were closed with two pieces of Blu Tack. Another multitest glass slide was pressed on top of the Blu Tack, until a perfect cylindrical shape was formed by the solution droplets from the side view (Figure 2.10). This sandwich structure prevented the refraction of the fluorescent light by the spherical shape of the solution droplet.

(a)



(b)

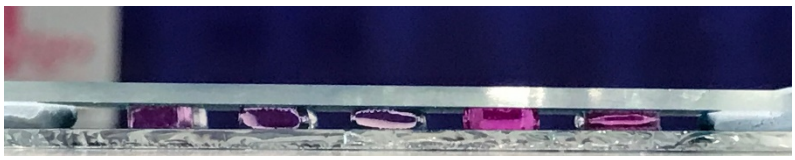


Figure 2.10– (a) The top view and (b) side view of the sandwich structure for mcRFP solution activity determination under BiognostiX reader.

A Matlab script was developed to calculate the total intensity of protein solution and protein fibre. The images obtained for the fibres and solution were cropped by imageJ so that only the area of each sample was included in each image for further processing with Matlab (Figure 2.11). The cropped images reduced the errors coming from the background. The red colour was then seperated from the green and the blue colour for each image. The intensity of the colour

was then represented by a number between 0 and 255, and a histogram was generated showing the pixel distribution at different intensities (Figure 2.12). The highest intensity value found from the histogram of the 0 μM solution sample (Figure 2.12 (a)) was used as the threshold of the background and was subtracted in each sample. The total intensity is calculated by adding up the intensity at each pixel. Figure 2.12 (b) and (c) show the histogram of the 8.95 μM mcRFP solution and the mcRFP fibre, respectively. From Figure 2.12 (a), the threshold of the background is set as 8. The total number of pixels whose intensity are higher than the threshold for the mcRFP solution (Figure 2.12 (b)) is 41830, whereas the total number of pixels whose intensities are higher than the threshold for the mcRFP fibre (Figure 2.12 (c)) is 2867.

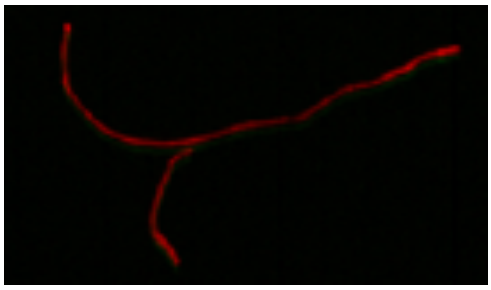
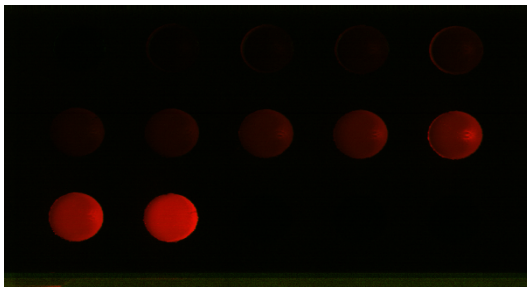
(a)  (b) 

Figure 2.11– Examples of BiognostiX images of (a) mcRFP protein solution assay, and (b) mcRFP protein fibre. (b) was cropped around the shape of the fibre to minimize background influence.

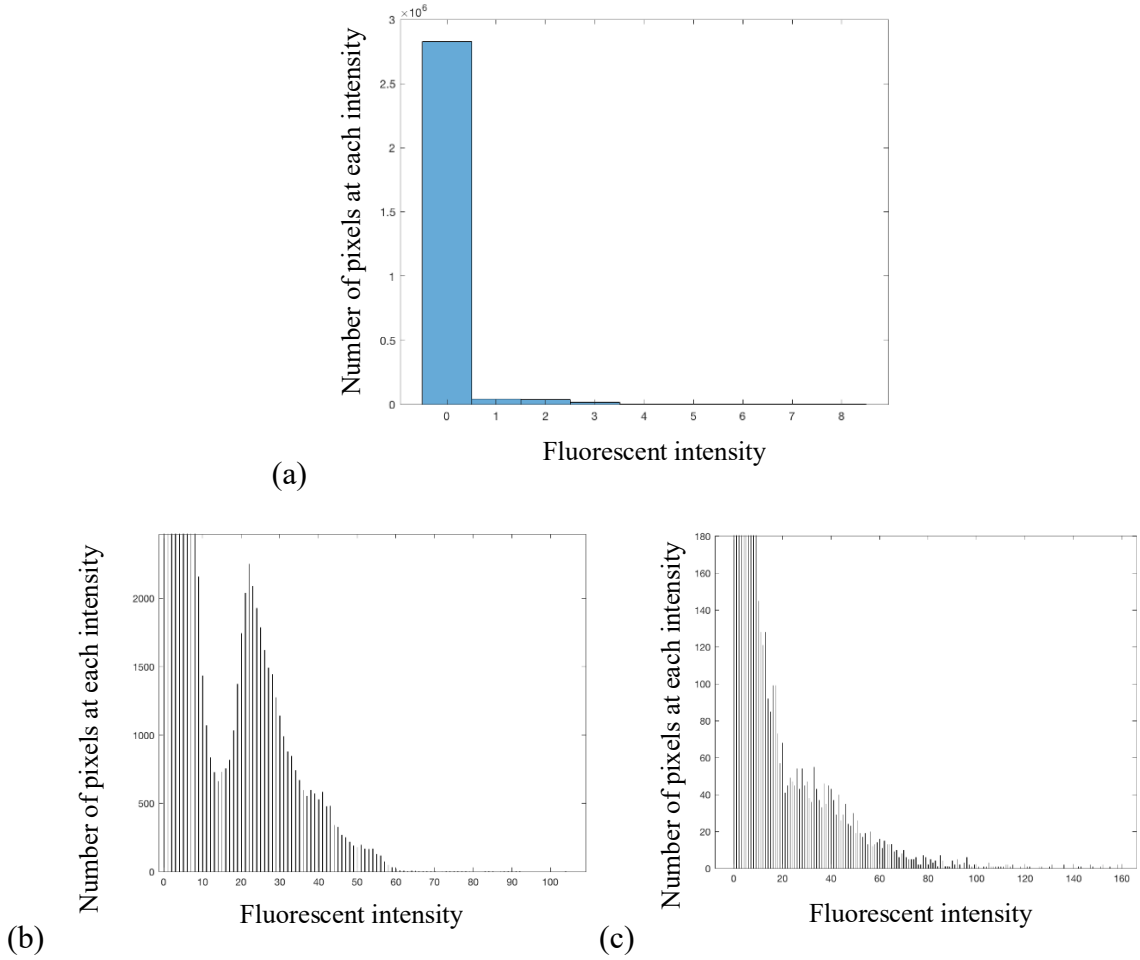


Figure 2.12– Histograms of (a) background, (b) 8.95 µM mcRFP solution, and (c) mcRFP fibre.

2.9.2.2 Spectroscopic Methods for Luciferase Protein Fibres

The standard curves for measuring the bioluminescent activity in Luciferase protein fibres were taken with the corresponding Luciferase protein solutions. The flash intensity of Luc-mcRFP and Luc-mcRFP-Q77 was recorded in 90 µL of 1 mM ATP and 200 µM D-LH_2 in 1xTEM solution using the Luminometer with 10 µL protein. The amounts of the proteins varied from 2.5 – 20 fmoles for LR, and 2.5 – 12.5 fmoles for LRQ. The corresponding fibres were put at the bottom centre in each well of the black 96 well plate. 10 µL of storage buffer was added in each fibre sample before the injection of 90 µL of substrate solution at the same concentration for the protein solutions. The bioluminescent intensity was recorded over time. Same procedure

was followed with Luc and Luc-Q77 protein fibres. And the maximum bioluminescent intensity was plotted against the calculated fibre volume.

ATP assay was tested with Luc-mcRFP solution and Luc-mcRFP fibre. A piece of fibre or 1 pmole solution was added to a black 96 well plate. D-LH_2 was added to a final concentration of 200 μM. 1 mM ATP was injected continuously to the reaction system to achieve different final amount of ATP. The bioluminescence intensity was recorded over time at each amount of ATP at 800 V for LR fibre, and 600 V for LR solution.

Similarly, the absorption spectra of protein fibres with mcRFP in the constructs were obtained with the UV-vis LAMBDA 16 spectrophotometer. Transparent 96 well plates were used to carry 200 μL sample solution in each well. Protein fibres were carefully stuck at the bottom centre in each well. 200 μL phosphate buffer was added to the fibre sample. The absorption was recorded from 400 nm to 700 nm. The scanning steps were set as 1 nm.

Chapter 3 Functional Recombinant Proteins

3.1 Introduction

The aim of this Chapter is to produce the recombinant proteins for further assembly experiments, and to compare the activity of different protein constructs and explain the structural insights to this change of activity. The activity change upon fusion is the basis for further protein engineering combined with the protein self-assembly and interfacial assembly. Two structurally distinct proteins were selected as the functional partners: monomeric red fluorescent protein (mcRFP), which is relatively small in size (~29 kDa) and possesses a stable β-barrel, and firefly luciferase (Luc), which is a larger protein (~63 kDa) with the active site located between two mobile domains. Fusion of the functional partners with the polyglutamine (Q77) tag initiates the formation of inclusion body during protein production but could potentially facilitate the co-assembly of the protein molecules into multi-molecular structures.

This chapter compares the fusion effect on mcRFP by either Luc, Q77 or both, and in turn compares the changes of Luc fused with either mcRFP, Q77, or both. Each fusion has different effect on the parent protein. Different characterisation methods will be discussed to carry out a full image of protein structural and activity change, and the relationship between the two. With mcRFP, the maturation process of the chromophore was a focus, together with its fluorescence properties revealing the folding of the β-barrel. Apart from the bioluminescent intensity, substrate binding kinetics are essential parameters indicating the folding of these two-domain luciferase structures. The choice of these two proteins is also providing an opportunity for a Bioluminescent Resonance Energy Transfer (BRET) due to the overlap between the emission of luciferase and the excitation of mcRFP.

3.2 Results and Discussions

3.2.1 Polyglutamine (Q77) Tagged Functional Proteins

To investigate the process of protein assembly and pursue the polymeric protein products, the monomeric precursors are first produced. Six recombinant proteins (mcRFP, mcRFP-Q77, Luc, Luc-Q77, Luc-mcRFP, and Luc-mcRFP-Q77) were expressed by *E. coli*, before being harvested and purified *in vitro* using the affinity between the fused 6xHistidine tag and a nickel column. For mcRFP-Q77, different expression conditions have been attempted with a range of incubation times after induction with IPTG. Figure 3.1 shows the results from different expression conditions on an 8% SDS-PAGE gel. Lanes 2-4 compare the purified mcRFP-Q77 with pure mcRFP and recombinant mcRFP genetically fused with a honeybee silk protein (HSP), which shows the relative sizes of different mcRFP fused proteins. Lanes 5-10 are crude mcRFP-Q77 lysate (lanes 5-6), the second wash of the crude mcRFP-Q77 lysate (lanes 7-8), and the purified mcRFP-Q77 (lanes 9-10) under different expression conditions. The red circled bands in both lane 2 and lane 9 possess sizes around 41 kDa. The sequence results, however, indicate the correct identity in lane 2 only with 63% total amino acid match of the theoretical mcRFP-Q77 sequence. This low match is consistent with the issues in detecting the large $[polyQ]_R$ peptide by normal mass spectroscopy, and therefore a high-mass examination was used. The peak suggested a polyQ region of 75 repeats (Appendix A.2).

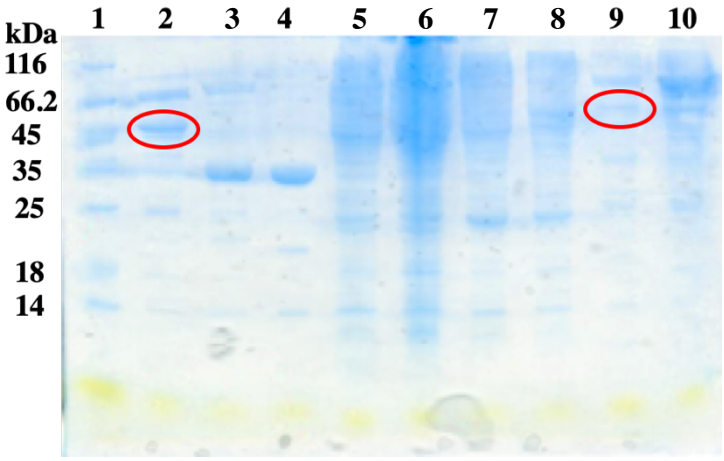


Figure 3.1– 8% SDS-PAGE gel of different expression conditions of mcRFP-Q77: purified mcRFP-Q77 (lane 2), crude lysate (lane 5), second wash of the crude lysate (lane 7) expressed for 5 hours at 37 °C followed by 14 hours at 17 °C; purified mcRFP-Q77 (lane 9), crude lysate (lane 6), second wash of the crude lysate (lane 8) expressed for 5 hours at 37 °C; and purified mcRFP-Q77 expressed for 5 hours at 37 °C followed by 12 hours at 17 °C (lane 10). Lane 3 and Lane 4 compares the size of the purified mcRFP-HSP and mcRFP, respectively, both expressed for 5 hours at 37°C. Lane 1 is the unstained protein marker.

Comparing with mcRFP, mcRFP-Q77 requires longer expression time for enough protein to be purified and identified. It has been reported that the extended polyQ over 42 glutamine repeats induces cell toxicity when expressed in *E.coli*.[214] As is discussed in Chapter 1, this toxicity is associated with the large intracellular aggregates formed from the monomeric protein molecules. Therefore, the soluble fraction left for native state purification is limited and may only be adequate for downstream purification after a longer time of accumulation during expression. The production protocol of recombinant Luc proteins was developed by Jaward Ur Rehman, and the full DNA sequence is shown in Appendix A.1.[206] Both the polyQ tagged Luc and Luc-mcRFP achieved the highest yield with 6 hours of expression, whereas 8 hours are needed for the maximal production of those without polyQ tag. Such difference in expression condition might indicate a lower cell toxicity effect for Luc fusion with Q77, due to the fact that polyQ induced toxicity and IB formation is fusion partner dependent.[83] Table 3.1 compares the maximum yield achieved for all of the protein constructs. mcRFP, Luc, and Luc-mcRFP are soluble proteins without any IB formation, with 68, 32 and 55 mg protein produced from

the soluble fraction of 1 L cell culture, respectively. The final yield of the IB forming Q77 tagged proteins were much lower, and the amount of proteins purified from the soluble fraction were limited to be 3, 4 and 8 mg/L of cell culture for mcRFP-Q77, Luc-Q77, and Luc-mcRFP-Q77, respectively. Figure 3.2 showed bands that were consistent with the expected identity of all the six constructs produced.

Table 3.1 – Maximum yield achieved for each protein construct through expression and extraction from *E. coli*.

Protein	mcRFP	mcRFP-Q77	Luc	Luc-Q77	Luc-mcRFP	Luc-mcRFP-Q77
Yield (mg/L)	68	3	32	4	55	8

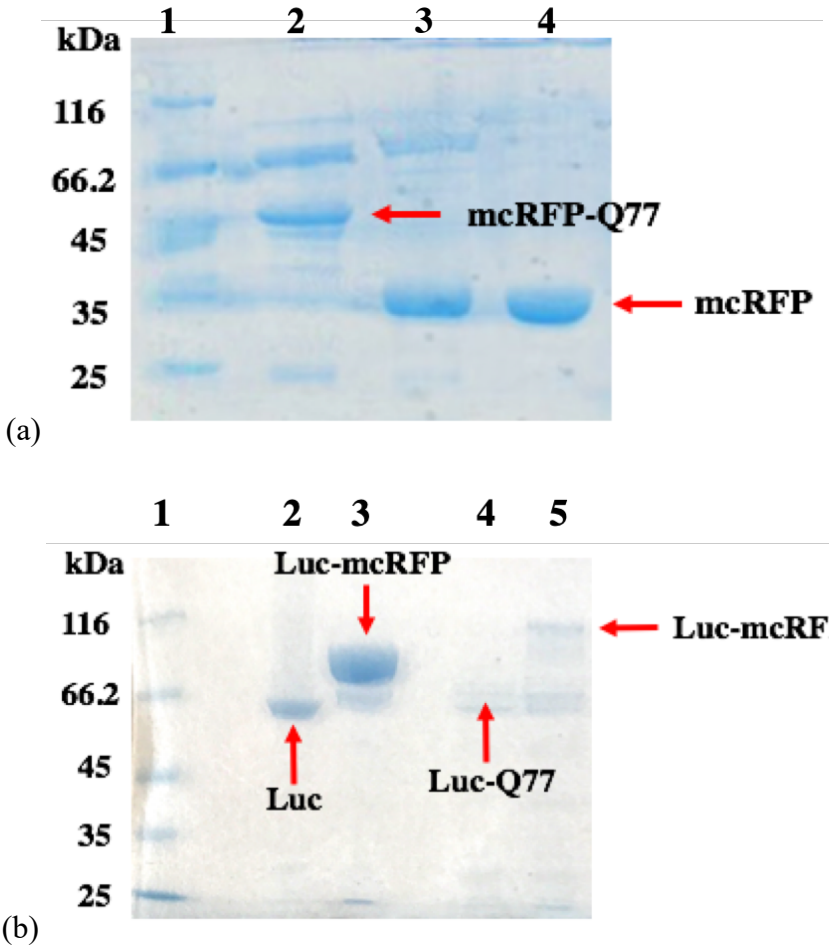
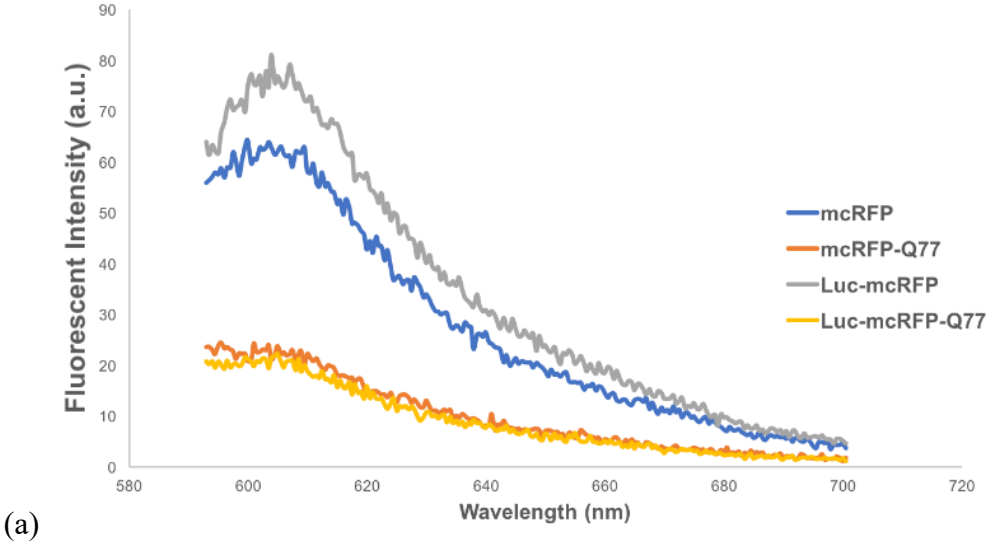


Figure 3.2– SDS-PAGE gel of purified recombinant protein. (a) mcRFP constructs on 8% SDS-PAGE gel. Lane 1: unstained protein marker; lane 2: mcRFP-Q77 (41 kDa); Lane 3: mcRFP-HSP (66 kDa); Lane 4: mcRFP (28 kDa). (b) Luciferase constructs on 6% SDS-PAGE gel. Lane 1: unstained protein marker; Lane 2: Luciferase (61 kDa); Lane 3: Luciferase-mcRFP (89 kDa); Lane 4: Luciferase-Q77 (75 kDa); Lane 5: Luciferase-mcRFP-Q77 (103 kDa).

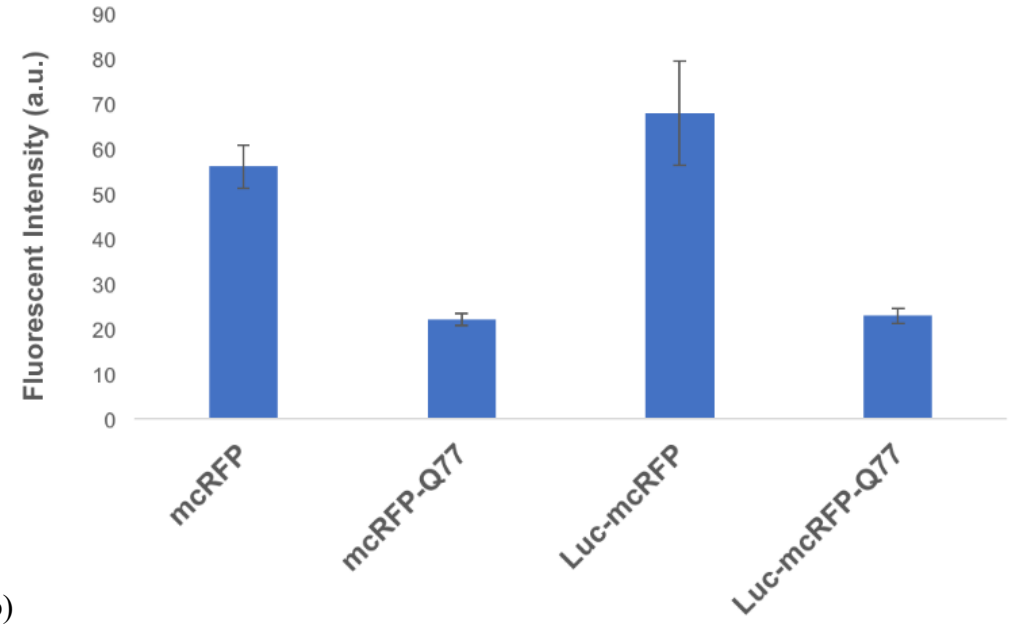
3.2.2 Fusion Effect on Red Fluorescent Protein

3.2.2.1 Fluorescence Characterisation of Fusion Proteins

Results show that the functional proteins lose both partial fluorescence and absorbance upon fusion with Q77 (Figure 3.3 and Figure 3.4). The fluorescent intensity is one of the important characteristics implying the folding of mcRFP. The tripeptide chromophore in mcRFP can only form if the protein scaffold is folded in a manner, which brings the tripeptide into correct alignment, i.e. cis-coplanar conformation. Upon the C-terminal fusion with Q77, mcRFP retains 40% of its original fluorescent activity. In contrast, the fluorescent activity enhances by 25% when the N terminal of mcRFP was fused with Luciferase. Such increase in fluorescence emission was eliminated when Q77 was further fused at the C terminal of the Luc-mcRFP, giving the same value of fluorescent intensity for both mcRFP-Q77 and Luc-mcRFP-Q77 (Figure 3.3). These fluorescent changes are indicative of some deformation in the mcRFP scaffold, especially the negative effect from the C-terminal Q77.

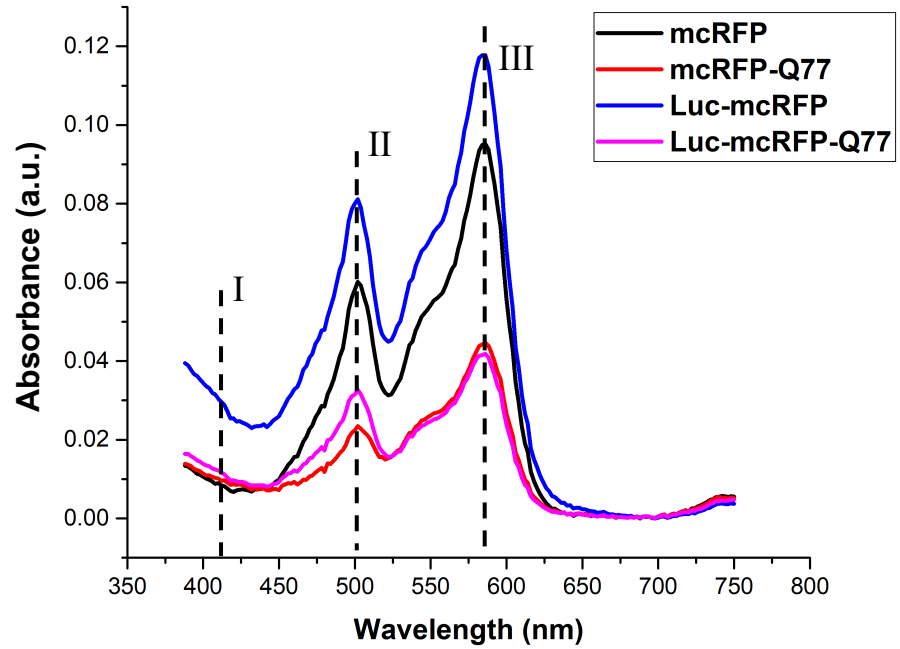


(a)

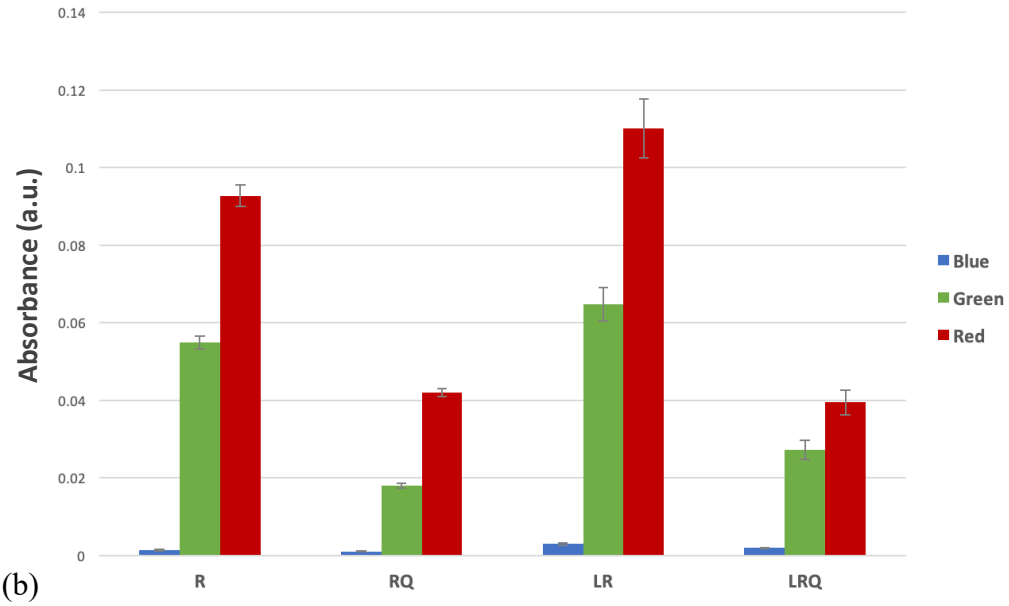


(b)

Figure 3.3– Fluorescence of different recombinant mcRFP in solution. (a) Fluorescent spectrum of recombinant mcRFP in solution. (b) Comparing fluorescent intensity of different recombinant mcRFP in solution. The concentration of all protein solutions is 5 μM.



(a)



(b)

Figure 3.4– Absorption of recombinant mcRFP in solution. (a) Absorption spectrum of each recombinant mcRFP in solution. The absorption peaks for mcRFP are at 408 nm (I: blue), 502 nm (II: green), and 586 nm (III: red) (b) Comparing absorbance at each peak wavelength of each recombinant mcRFP in solution. The concentration of all protein solutions is 5 μM.

Similar trend was observed with the absorption spectra. The absolute value of red absorption decreased with Q77 tag, while increased with the N-terminal Luc fusion (Figure 3.4 (a)). The

absorption is an indication of total chromophore formation. As is discussed in Chapter 1, the latest proposed mechanism of DsRed chromophore maturation undergoes the "TagBFP"-like intermediate (Figure 3.5). The branching point of the chromophore occurs after cyclisation, followed by either a dehydration/oxidation pathway through a "TagBFP"-like intermediate towards the matured red chromophore; or an oxidation/dehydration pathway through a GFP-like intermediate towards the red chromophore with an extra dehydrogenation (alkane oxidation) process. Figure 3.4 (b) compares the relative absorptions from the blue intermediate, the green intermediate, and the red chromophore in different fusion constructs. It is previously reported that all the three chromophores possess very close extinction coefficient of around 44,000 $M^{-1}cm^{-1}$.[91] Therefore, the percentage of absorption from each chromophore could be calculated from the total chromophore absorption (Table 3.2).

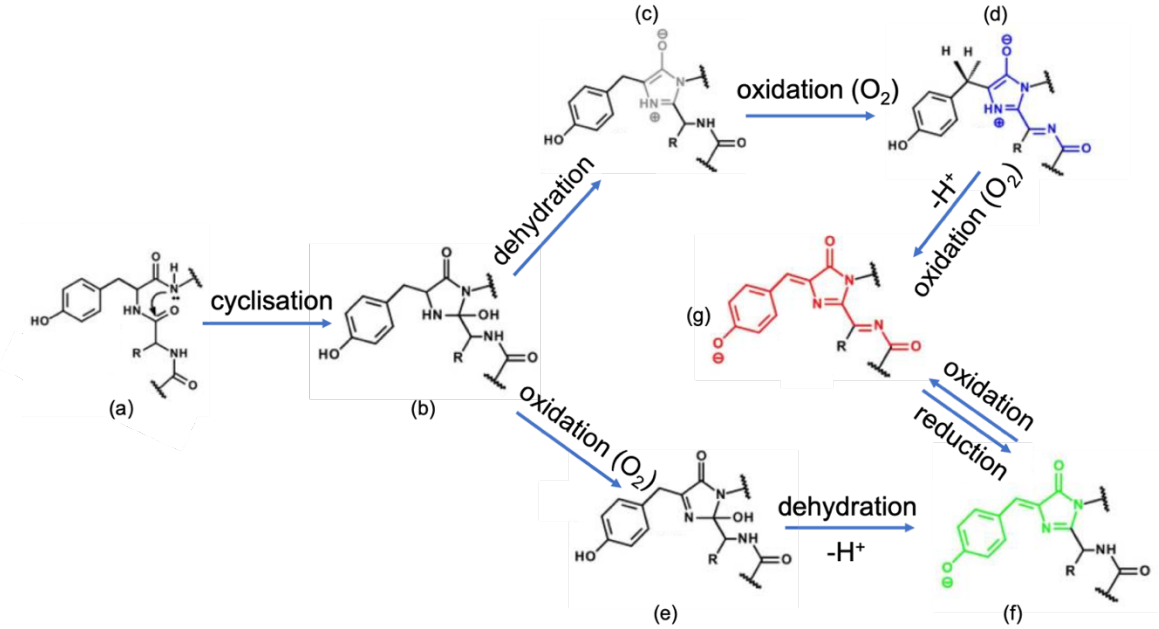


Figure 3.5– Chromophore maturation pathway of DsRed described by the "via-TagBFP" model. (Reproduced from Subach et al)[92]

Table 3.2 – Chromophore absorption ratio of recombinant mcRFP in solution.

Protein	mcRFP	mcRFP-Q77	Luc-mcRFP	Luc-mcRFP-Q77
Blue intermediate / total chromophore	0.9% ± 0.16%	1.8% ± 0.32%	1.7% ± 0.24%	2.9% ± 0.38%
Green intermediate / total chromophore	36.9% ± 0.08%	29.5% ± 0.38%	36.4% ± 0.09%	39.7% ± 0.38%
Red end product / total chromophore	62.2% ± 0.08%	68.7% ± 0.26%	61.9% ± 0.18%	57.4% ± 0.22%

From Table 3.2, a decreased proportion of the GFP-like intermediate was observed with mcRFP-Q77 comparing to mcRFP. Together with the increased proportion of the blue chromophore, the result might be explained as a decreased protonation during maturation. The increased GFP-like intermediate might indicate the misfolding of the β-can surrounding the chromophore and reflects shielding/deshielding by the polyQ β sheets. Intramolecular hydrogen bonding in polyQ causes it to form a β-sheet structure associated with Q, but this may also 'recruit' and potentially divert the native intramolecular bonding involved with the β-can. If this involves the native hydrogen bond network, it may potentially prevent proton abstraction. Furthermore, the higher proportion of blue chromophore represents the failure of the deprotonation step to the red chromophore, suggesting that the interactions with the polyQ is detrimental. The red end product, however, increases for mcRFP-Q77. This is explained by the fact that the process of the alkane oxidation (Figure 3.5 f→g) is in competition with the deprotonation steps (Figure 3.5 e→f).[91] The less deprotonation results in a favourable alkane oxidation process, which further decreases the percentage of the green intermediate and causes the increase in the percentage of the red chromophore.

With Luc fusion, the green intermediate remained in the similar range with the pure mcRFP. The blue chromophore, however, takes up a higher proportion. Higher percentages of both the green and the blue chromophores were obtained with dual fusions of Luc and Q77 to mcRFP at opposite terminals, with a decreased red chromophore. Both structures prove that the fusion of Luc has a larger effect on the prevention of the last O_2-mediated oxidation/deprotonation step from the blue intermediate towards the red end product, which might be due to a stable Luc-fused structure with the oxidised blue chromophore. Instead of the hydrogen bonding network associated with the β-sheet polyQ, the larger proportion of green chromophore in Luc-mcRFP-Q77 is indicative of a preferred dehydration against alkane oxidation. This might be resulted from the more rigid mcRFP structure with the double fusions of Luc and Q77 at opposite terminals.

Despite the decrease in the intensity of the peak, no difference in the shape of the fluorescence emission and the absorbance spectra was found (Figure 3.3 (a) and Figure 3.4 (a)). The absorbance peaks remained at 502 nm for the green chromophore, and at 586 nm for the red chromophore, regardless of the fusion. Excitation at 586 nm gives an emission peak maximum at 607 nm, which is also indicative of a correct chromophore conjugation system with no further chromophore post modifications.

Quantum yield is a measurement of the efficiency of light production in a fluorescent system. Table 3.3 lists the quantum yield (QY) of all the mcRFP fused proteins. Although a decreased absorption and emission was observed with mcRFP-Q77, the QY remained comparable with mcRFP. The luciferase fusion either with or without the Q77 tag, increased the QY of the mcRFP chromophore. It is widely accepted, that in fluorescent proteins, the QY value is largely related to the rigidity of the chromophore.[215,216] Therefore, these data indicate that the chromophore is rigidified with the Luc fusion, which might contribute to the less dehydration process occurring for the formation of the red chromophore. The high quantum yield of Luc-mcRFP also explained the high fluorescence obtained for this specific recombinant protein.

Table 3.3 – Fluorescent Quantum Yield (QY) and lifetime of recombinant mcRFP in solution.

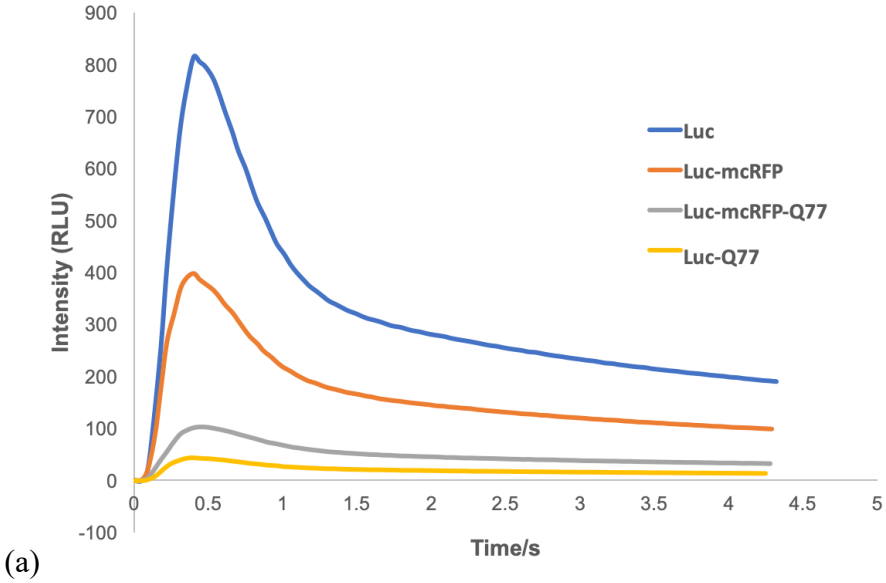
Protein		mcRFP	mcRFP-Q77	Luc-mcRFP	Luc-mcRFP-Q77
QY		0.165	0.167	0.221	0.180
Lifetime (ns)	Blue	0.96	0.12	0.01	0.02
	Green	2.22	1.79	1.12	0.64
	Red	4.25	2.37	1.98	1.92

The other important character for the chromophore in the fluorescent protein is the fluorescent lifetime (τ). Stepanenko et al[217] have reported monomeric RFP with a fluorescent lifetime of 1.75 ns measured at 610nm, following excitation at 365nm. The lifetime increases to >3 ns for the oligomeric red fluorescent proteins, where a decrease in solvent accessibility to the fluorophore is proposed to increase the lifetime. In contrast, the average fluorescent lifetime decreases in the immature red fluorescent protein due to the impact on the red fluorescence by a green fluorescent chromophore (τ <0.5 ns) that may also be formed. Peter et al[218] suggest that, after the maturation phase, monomeric red fluorescent protein does not produce significant green fluorescence, so the lifetime can be fitted to a single exponential with τ~2 ns. The mcRFP fused proteins in this thesis absorb blue, green and red light at 408, 502 and 586 nm, respectively, and are fitted with 3 exponential decay corresponding to the lifetime of blue, green and red chromophores. It is noticed that the lifetime of each chromophore decreased when mcRFP was fused with either Q77, Luc, or both. A decrease of 44% in the lifetime of the mcRFP red chromophore was observed when it was fused with Q77 (Table 3.3). Similarly, a decrease of 53% and 55% was observed with Luc-mcRFP and Luc-mcRFP-Q77, respectively (Table 3.3). These data suggested an increased solvent accessibility to the chromophore for both Q77 and Luc fused mcRFP. Such a decrease in lifetime indicates a more rigid chromophore with increased solvent accessibility, due to the partial unfolding of the chromophore. The increased rigidity matched the previous discussion of the absorption change

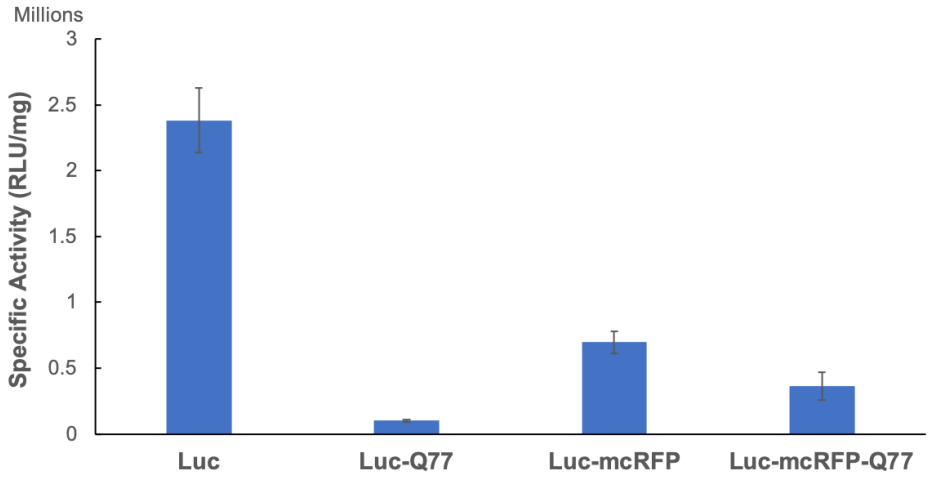
for Luc-mcRFP-Q77, which results in the less favoured alkane oxidation process from the green intermediate towards the red chromophore. Whereas for mcRFP-Q77 and Luc-mcRFP, the relative absorptions of the three chromophores show that a dominate effect of the less deprotonation processes dominate over the effect from the rigidified mcRFP structure. This could also be concluded combined with the less decreased lifetime of the single fused proteins comparing to the dual fused Luc-mcRFP-Q77. However, the QY was increased for all fusion constructs. This could make sense, as the QY value is indicative of the radiation process of the photons, whereas the non-radiative energy decay is also taken into considerations in the lifetime measurements. Therefore, the lifetime value is mainly revealing an increased solvent accessibility, despite the chromophore being rigidified by its fusion partners.

3.2.2.2 Fusion Effect on Luciferase

Similar to mcRFP, the Q77 tagged luciferase experienced a loss in its catalytic activity, with only 8% activity retention. In comparison, 34% activity was retained when mcRFP was used as a fusion partner. Further fusion of the recombinant Luc-mcRFP (LR) by the Q77 tag showed 19% retention of the total Luc activity (Figure 3.6 (b)). Figure 3.6 (a) illustrates the time course of the bioluminescent light emission. All constructs emit light with a maximum peak after 0.5 seconds, indicating the speed of the product formation remained the same, and the emission of the light is inhibited by the product subsequently, resulting in a decrease in the bioluminescent intensity after the flash height.



(a)



(b)

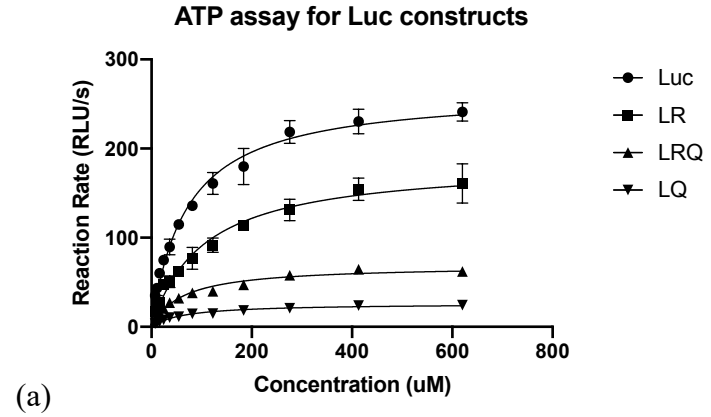
Figure 3.6– (a) Bioluminescence kinetics of recombinant Luc proteins in solution at a concentration of 50 pM. (b) Specific bioluminescent activity in recombinant Luc proteins in solution. (PMT=800 V)

The kinetic constants, K_m, of different luciferase enzymatic reactions were obtained by fitting the experimental data to the Michaelis-Menten equation using the Prism software (Figure 3.7 and Table 3.4). The K_m value indicates the affinity of the enzyme towards a certain substrate. The smaller the K_m value, the higher the affinity is. Comparing to the literature reported K_m values for wild type luciferase, which are 13.9 ± 2.4 μM for ATP, and 61.7 ± 3.2 μM for D-LH_2,[101] the in-house produced recombinant Luc showed a decreased affinity to ATP and a slightly increased affinity towards D-LH_2. The turnover number, k_{cat}, was determined by the

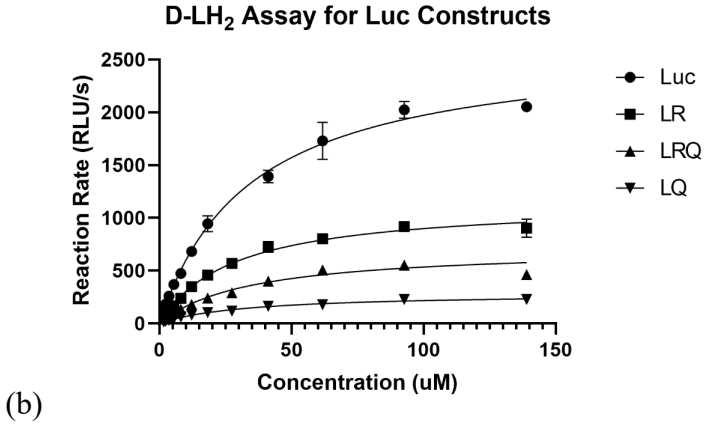
V_{max} value. Different voltages were used for the ATP assay and the D-LH_2 assay, so that the bioluminescent intensity at each concentration are within the limit of detection of the Luminometer. The relative trend of the V_{max} values are the same for each protein constructs, even though the absolute numbers are different. Therefore, the k_{cat} number was calculated from the V_{max} value of ATP to represent the relative turnover rate of each construct (Table 3.4). Despite a large decrease in both the specific activity (Figure 3.6) and the k_{cat} number (Table 3.4), when Luc was fused with the polyQ tag, the substrate affinities for both ATP and D-LH_2 were similar, which suggests that the active sites for the substrates have not been compromised. The mcRFP fusion was exceptional. Luc-mcRFP maintained more specific activity (34%) and experienced the least decrease in the turnover rate. The substrate affinity for ATP, however, is largely decreased, with K_M value increased by 48%. In contrast, the affinity towards D-LH_2, increased, with K_M value decreased by 30%. Further polyQ fusion to Luc-mcRFP showed similar enzyme kinetics with both Luc and Luc-Q77.

Table 3.4 – Kinetic parameters of recombinant luciferase proteins.

	K_M • ATP (μM)	K_M • D-LH_2 (μM)	k_{cat} • ATP ($\times 10^9$ RLU/mol • s)
Luc	70 ± 5	34 ± 4	10.6
Luc-mcRFP	104 ± 11	28 ± 2	7.4
Luc-mcRFP-Q77	63 ± 5	34 ± 43	2.8
Luc-Q77	65 ± 6	33 ± 3	1.0



(a)



(b)

Figure 3.7– Fitted data of (a) ATP assay (PMT voltage 550 V) and (b) D-LH_2 assay for different Luc constructs with 2.5 μM of each enzyme (PMT voltage 400 V).

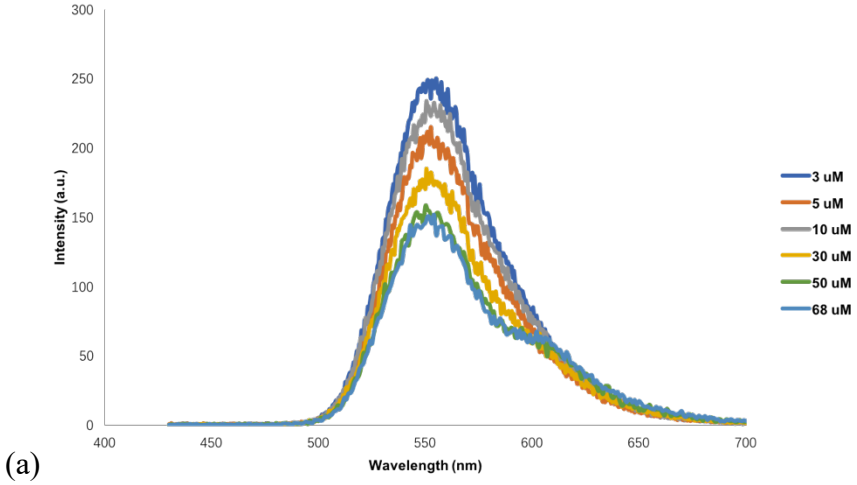
Such kinetic changes are indications of some conformational changes after fusion. It was reported by Zako *et al*, that a deletion of the small C terminal domain would result in a much slower light emission, which took 4 minutes to achieve the flash height.[219] In addition, the two domains in luciferase have been studied to follow an "open-close-open" pathway during the light emission reaction, whereas the two domains in a reported mutant luciferase were found to remain "open", which results in the emission of a red coloured light.[220] Therefore, it can be deduced from the results, that the two domains of luciferase are still intact, and the two domains follow the "open-close-open" path during the light emission reaction, with either mcRFP fusion, or polyQ fusion, or mcRFP-Q77 fusion, given the fact that the colour emissions catalysed by these recombinant luciferases remained at yellow-green at 550 nm, and that the flash height was still achieved at 0.5 seconds.

Previous report showed that the most luciferase active site is located at the N terminal domain, with only the lysine at K529 position in the C domain responsible for luciferase activity. Zako *et al* further demonstrated that the deletion of C domain slightly decreases the affinity towards D-LH_2, but dramatically reduces ATP affinity. This indicates the role of D-LH_2 orientation for the C domain, but a much more important role of ATP binding.[219] Therefore, the exceptional behaviour of LR can be explained as a larger conformational change at the C domain, compared to LQ and LRQ, suggested by the increased ATP binding affinity.

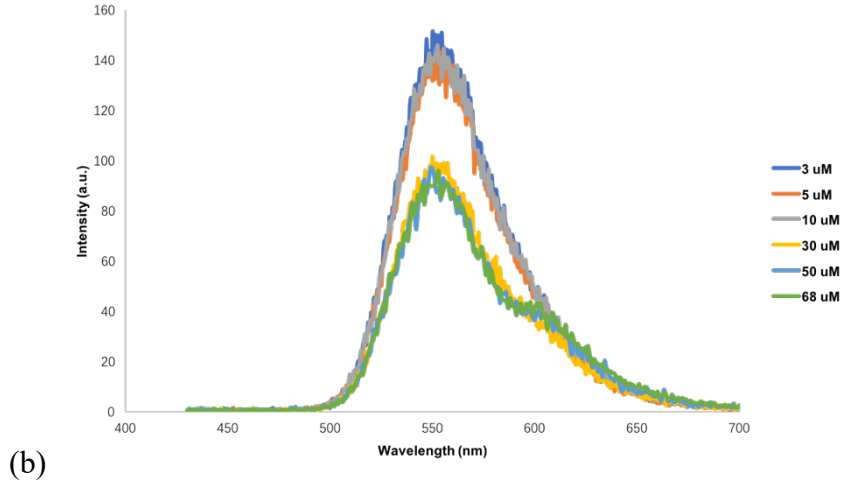
Another important character worth investigation upon fusion of the two luminescent proteins is the Bioluminescent Resonance Energy Transfer (BRET). BRET happens when a bioluminescent donor comes close to a fluorescent acceptor in the presence of a substrate. In theory, the distance between the donor and the acceptor required for BRET to occur should be less than 10 nm.[221] In this case, the fluorescent acceptor emits fluorescent light without the requirements of any external light. The two protein models used in this study possess the possibility of BRET to occur, as the yellow-green light (550 nm) emitted from the Luc happens to be in the range of the exciting wavelength for mcRFP fluorescence (maximum excitation at 586 nm). When Luc catalyses the oxidation of D-LH_2 in the presence of ATP and Mg^{2+}, the bioluminescent light excites the nearby mcRFP molecules to emit light at 610 nm.

A study of intermolecular BRET showed this transfer. From Figure 3.8 (a), such intermolecular BRET would only occur when mcRFP concentration is increased to 30 μM in 1 μM Luc. Lower concentration of mcRFP may result in larger molecular distances between Luc and mcRFP in the solution for an efficient BRET to occur. Increased BRET effect (fluorescent intensity at 610 nm) resulted in a decrease in the bioluminescent intensity from Luc at 550 nm. With longer reaction time, the bioluminescent intensity decreased, and the BRET intensity became concentration independent (Figure 3.8 (b)). However, during the time of the bioluminescent decay (e.g. from 22 s to 51 s), the BRET ratio (intensity at 610nm/ intensity at 550 nm) remained the same for the same reaction system. Such effects, however, did not happen in the fused proteins, Luc-mcRFP and Luc-mcRFP-Q77. This may be associated with orientation for the dipole-diploe coupling required to achieve BRET. Such dipole-dipole coupling is dependent on the geometry of the donor and acceptor proteins. The fusion rigidified the two

protein structures in a certain geometry, which preventing them from rotating freely as separate molecules.



(a)



(b)

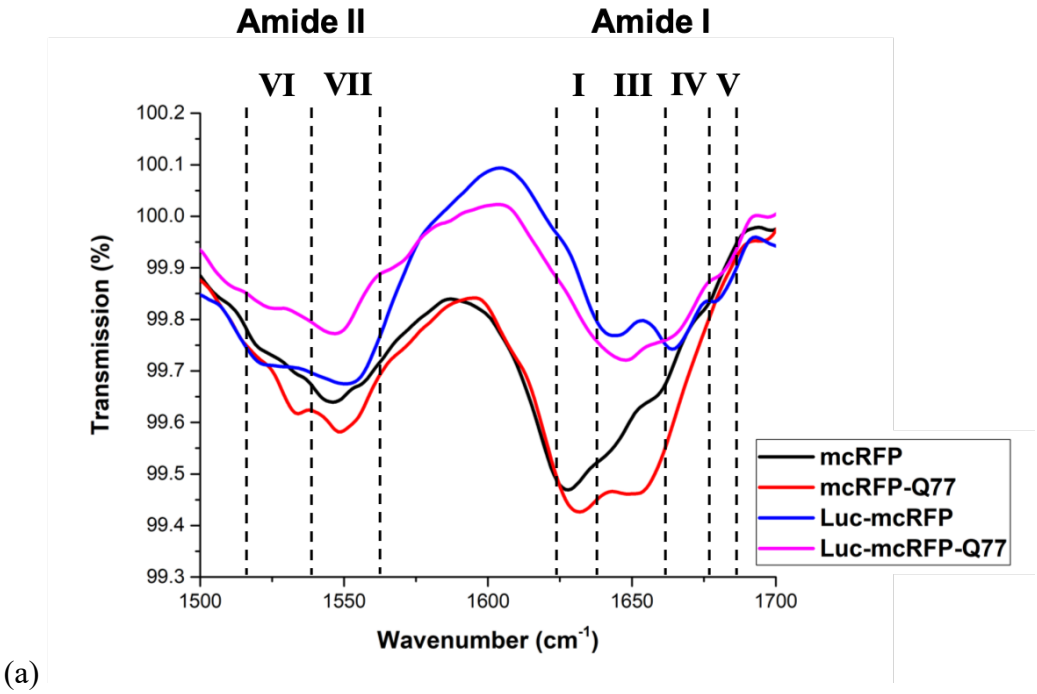
Figure 3.8– Intermolecular BRET profile between 1 μM free luciferase molecules and various concentrations of mcRFP molecules at (a) 22 seconds and (b) 51 seconds after substrates were added, and (c) inter/intra-molecular BRET spectrum comparing 10 μM Luc, 10 μM Luc-mcRFP, 10 μM Luc-mcRFP-Q77, and 10 μM free Luc with 10 μM free mcRFP at 22 seconds after substrates were added.

3.2.3 Structural Insights of Fusion Proteins

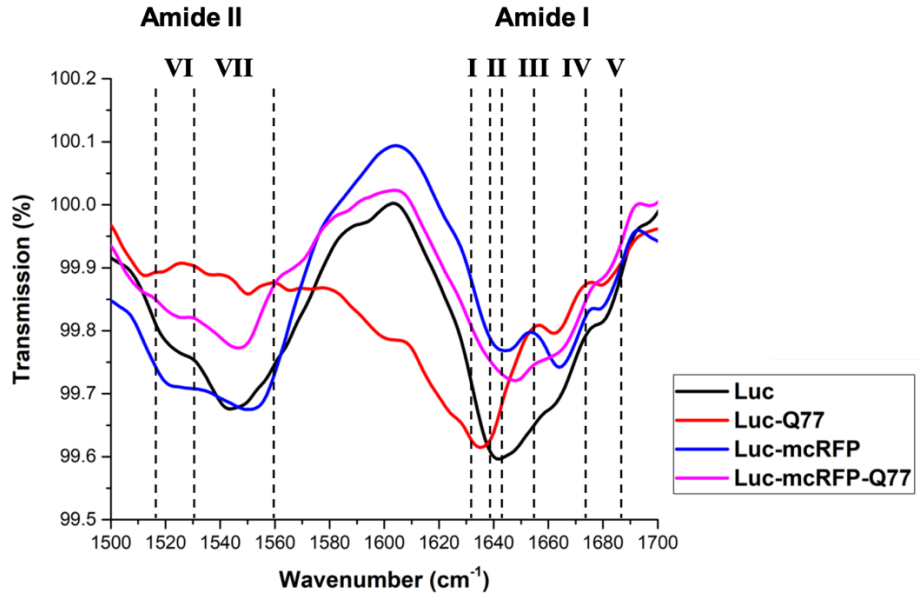
The change in activities of fusion proteins could be further explored by the determinations of protein secondary structures. In the protein FT-IR spectra, the wavenumber of interest ranges

from 1500 cm^{-1} to 1700 cm^{-1}. The absorption in this region consists of amide I and amide II vibration modes. The other amide bands are currently little use in protein secondary structure predictions, due to their complexity and high dependency on the force field, side chain, and hydrogen bonding.[222] The amide I vibration, absorbing around 1650 cm^{-1}, arises mainly from the C=O stretching. The amide I absorption is attributed from the internal coordinates of the backbone structures in proteins. The amide II mode mainly comes from the out-of-phase combination of the NH in plane bend and the CN stretching vibration. The correlation between frequency and protein secondary structure is less straightforward than amide I.[223]

α-helices give rise to absorption close to 1655 cm^{-1} in amide I region, which corresponds to an absorption around 1545 cm^{-1} in amide II region. The band position shifts down with increasing helix length, when the helix is bent in coiled coils, and when the helix is water exposed. Random coil structures absorb at similar range, and sometimes it is hard to distinguish between the two. In contrast, β-sheet structure exhibits a strong band near 1630 cm^{-1} and is often appeared splitting with weak absorptions at around 1685 cm^{-1}. The structure absorbs at 1530 cm^{-1} in amide II region. [223,224] From the amide I region of the FT-IR spectrum (Figure 3.9 (a)) and the peak assignments in Table 3.5, the polyQ tag largely increases the helical content of mcRFP, which is deduced from the appearance of a well-defined peak at 1652 cm^{-1} for mcRFP-Q77 spectrum, compared to the shoulder at 1658 cm^{-1} for mcRFP spectrum. The amide II region confirms the appearance of both helical and sheet structures at around 1545 cm^{-1} and 1532 cm^{-1}, respectively, in both proteins.



(a)



(b)

Figure 3.9– FTIR spectra of 2 μM recombinant protein solutions. Comparing (a) mcRFP recombinant constructs, and (b) Luc recombinant constructs in both amide I and amide II region. The dashed lines separated regions are representative of different protein secondary structures. I/VI: β-sheet; II/VII: random coil; III: α-helix; IV: β-turns; V: β-sheet Antiparallel.

Table 3.5 – Peak assignment of FT-IR spectra.

	β-sheet	Random Coil	α-helix	β-turns	β-sheet Antiparallel
mcRFP	1628, 1640 (s) 1532 (s)		1658 (s) 1545, 1552 (s)	1674 (s) 1528	
mcRFP-Q77	1632 1532		1652 1548		1697 (s)
Luc-mcRFP	 1520		1645 1552	1665	1678 (s)
Luc-mcRFP-Q77	 1523		1648 1548	1665 (s)	1680 (s)
Luc	 1523	1642	 1542	1662	1680 (s)
Luc-Q77	1635 1520		 1535, 1545	1662	1680 (s)

*(s) indicates shoulders or relatively small peaks.

With luciferase fused mcRFP, and the consequent Luc-mcRFP-Q77, the IR spectra showed the major absorption at a decreased wavenumber (~1645 cm^{-1}) for α-helix, with a large content of β-turns found in Luc-mcRFP. The decreased wavenumber of α-helix, which is caused by an elongated helical structure, is in the overlapping region with the random coil structure. Therefore, the structure is further confirmed in the amide II region, as well as the CD spectrum. Major peaks at around 1550 cm^{-1} in the Amide II region indicated the helical structure with noticeable sheet structure at around 1520 cm^{-1}. The spectra showed significant peak shifts compared to mcRFP (Figure 3.9 (a)), but similar peak positions as Luc with largely different peak ratios (Figure 3.9 (b)). Luc-Q77, in contrast, showed a different major peak, corresponding to a sheet structure. The absorption in the amide II region is weak, but it indicates the existence of both helical and sheet structure. With Luc, the major peak at 1642 cm^{-1} is assigned as a random coil structure.

Leading on from the identification of the secondary structures from the FT-IR spectroscopy, the detailed structures of the proteins were further determined by CD spectroscopy (Figure 3.10). The spectra were analysed by fitting the data using the CDSSTR program, which calculates the relative percentage of the secondary structure components (Figure 3.11).

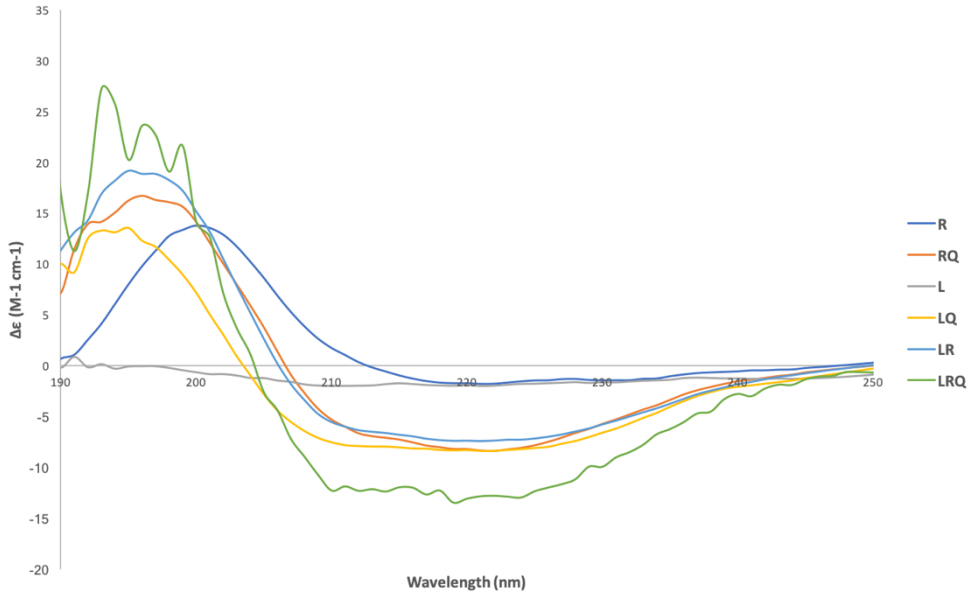


Figure 3.10– CD spectra of all recombinant proteins.

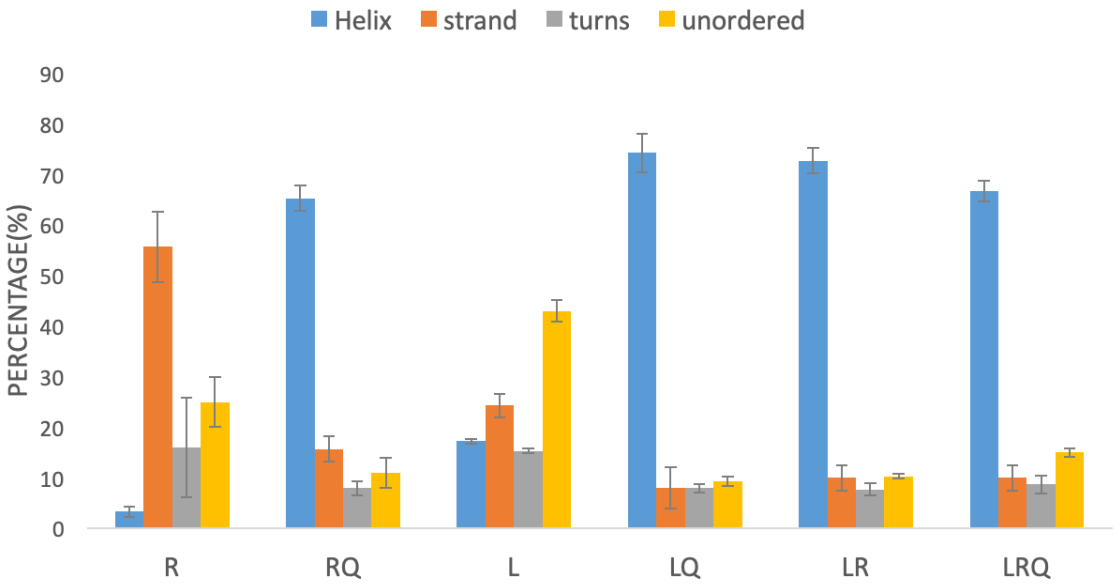
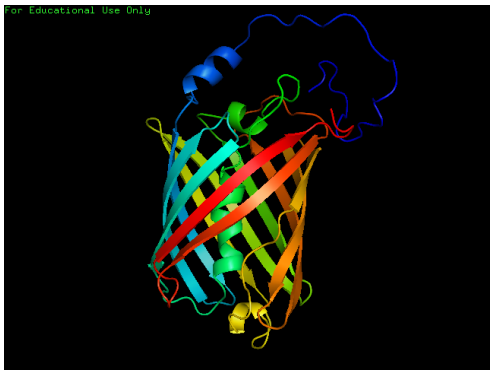


Figure 3.11– Content of secondary structures in each recombinant protein analysed from the CD spectra.

From the analysed CD spectra (Figure 3.11), the helical structure increased from 3% in mcRFP to 65% in mcRFP-Q77 while there was only 16% β-strand retained with Q77 fusion compared to the 55% in mcRFP. The structure was further predicted by the Robetta server and displayed with Pymol (Figure 3.12). Robetta is a free online server developed in Baker's lab, Department of Biochemistry, University of Washington.[225] Sequences submitted to the server are parsed into putative domains, and protein structure models are generated using either comparative modelling or *de novo* structure prediction methods.[226] For example, when the sequence of mcRFP-Q77, whose structure has never been crystallised and published, is submitted to the Robetta server, the sequence is parsed into mcRFP domain and Q77 domain according to the existing protein data base by comparative modelling. The two domains will be joint together with de novo structure prediction of the linker, and the final structure will be determined to the one with the least folding energy.

From Figure 3.12, there is an obvious structural change at the N-terminal loop of mcRFP comparing to mcRFP-Q77. For mcRFP (Figure 3.12 (a)), the N terminal starts from an unordered structure followed by a small helix structure before the β-barrel. This loosely unordered structure, however, wrapped into helical structure after the C terminal of mcRFP

was fused with Q77 (Figure 3.12 (b)). The change is explained by the spatial repulsion of the Q77 tag towards the unordered structure, forcing it to form a helical component. This structure prediction could explain the decrease of the unordered structure after Q77 fusion obtained by the CD spectrum. The calculated secondary structure indicates 11% of helical content and 49% of sheet structure in mcRFP, whereas the helix and sheet structure are 38% and 37%, respectively, in mcRFP-Q77 (Table 3.6). It can also be noticed that the predicted helical structure was lower, and the sheet structure was higher than that measured with the CD spectrum for mcRFP-Q77. Since it is suggested from the IR, the structural modelling and the CD spectroscopy, that Q77 adopted a helical structure, the Q77 tag should not contribute to any β-sheet component in mcRFP-Q77. If there is no structural change upon fusion of Q77, the 55% β-sheet structure in mcRFP should take up 37% ($\frac{28\ \text{kDa}}{41\ \text{kDa}} \times 55\%$) of the total mcRFP-Q77, which matches the modelled mcRFP-Q77 structure and is 21% higher than the sheet structure measured by CD spectroscopy. Therefore, it can be concluded that ~20% of the β-barrel unfolds in mcRFP when it is fused with Q77. This structural change could explain the decrease of the fluorescent lifetime in mcRFP-Q77 due to solvent exposure upon fusion with Q77.

(a) 

(b) 

Figure 3.12– Structures of (a) mcRFP and (b) mcRFP-Q77 displayed by Pymol, predicted by Robetta server.[225] The structures are oriented to a view with the best presentation of all domains in each protein.

Table 3.6 – Helical and sheet content calculated from the predicted protein structure.

	α-helix	β-sheet
mcRFP	11%	49%
mcRFP-Q77	39%	38%
Luc	35%	22%
mcRFP unfolded Luc-mcRFP	29%	20%
Luc unfolded Luc-mcRFP	26%	22%
Luc-mcRFP-Q77	34%	28%
Luc-Q77	44%	18%

The major peak of Luc observed in the FT-IR spectrum (Figure 3.9 (b)) matches the 46% unordered structure detected by the CD spectrum (Figure 3.11). All fused versions of Luc demonstrated nearly 70% helical structure in the CD spectra. The CD spectra matched the IR spectra of Luc-mcRFP and Luc-mcRFP-Q77, whereas the results for Luc-Q77 is controversial. This might be explained by a preferable hydrogen bond of the glutamine in Luc-Q77 with water. And it was reported that the protein interaction with water could lower the absorption frequency by 20-30 cm^{-1}. Therefore, the appeared β-sheet absorption in the FT-IR spectrum would be a hydrogen bonding interacted helical structure. More likely, this shift in the absorption wavenumber would be a further elongated helical structure from the fusion of the helical Q77 with Luc, which causes a decrease in the absorption frequency as discussed earlier.[223]

The predicted structure in Figure 3.13 showed little effect of Q77 on the structure of Luc, with 9% increase in the helical structure (Table 3.6). Similar calculation could be carried out here as with the mcRFP-Q77. The addition of helical Q77, should in theory increase the helical

content of Luc to 47% ($\frac{61\ \text{kDa}\times 35\%+14\ \text{kDa}}{61\text{kDa}+14\ \text{kDa}}$). This matches the calculated helical content of the predicted structure. However, the CD spectrum is showing a much larger percentage (75%) of helix, and the calculated sheet structure (18%) was determined to be ~10% from the CD analysis. Therefore, some deformation of the theoretical sheet structure in Luc occurred upon fusion with Q77, with 28% of helical increase. White spheres in Figure 3.13 show the K529 amino acid in the C domain of luciferase, which was reported to be responsible for D-LH_2 orientation, and ATP binding. The similar K_M values of Luc and Luc-Q77 indicates that there is not much differences in the environment of K529 for Luc and Luc-Q77.

Two different structures both resulted in the lowest folding energy when predicting the structure of Luc-mcRFP. One of them shows a largely denatured mcRFP structure with maintained Luc structure (Figure 3.13 (c)), whereas the other one retains the structure of mcRFP with unfolded Luc (Figure 3.13 (d)). According to the fluorescence data of mcRFP, the binding kinetics of Luc in this protein, and the largely decreased β-strand structure from the CD spectrum, it is deduced that both mcRFP and Luc experienced some conformational changes. It was previously discussed that the chromophore is rigidified with Luc fusion (higher QY) with increased solvent accessibility. Therefore, a partially unfolded mcRFP structure is consistent; the increased rigidity is supposed to come from the Luc interaction. With the largely decreased ATP binding affinity, the Luc structure should also have some conformational changes. The change in the environment of K529 (white spheres in Figure 3.13 (d)) supports the decrease in ATP binding affinity. In either case, it shows that the C-terminal fusion of Luc with mcRFP causes interference between the structures of both two proteins. This fusion has a larger effect on the C domain of Luc, which is highly related to ATP binding affinity. In contrast, the predicted structure shows that further fusion of Luc-mcRFP with Q77 protects the structure of C domain of Luc (Figure 3.13 (e)). A spatial protection effect of both the mcRFP and Luc was found by the insertion of the helical Q77. The experimental CD spectrum illustrates slightly less deformation, with less helical structure formed and more random coil structure retained, comparing with Luc (Figure 3.11 (b)). This structure is supported by the unchanged binding affinity (similar K_M values) of Luc-mcRFP-Q77 and the native Luc.

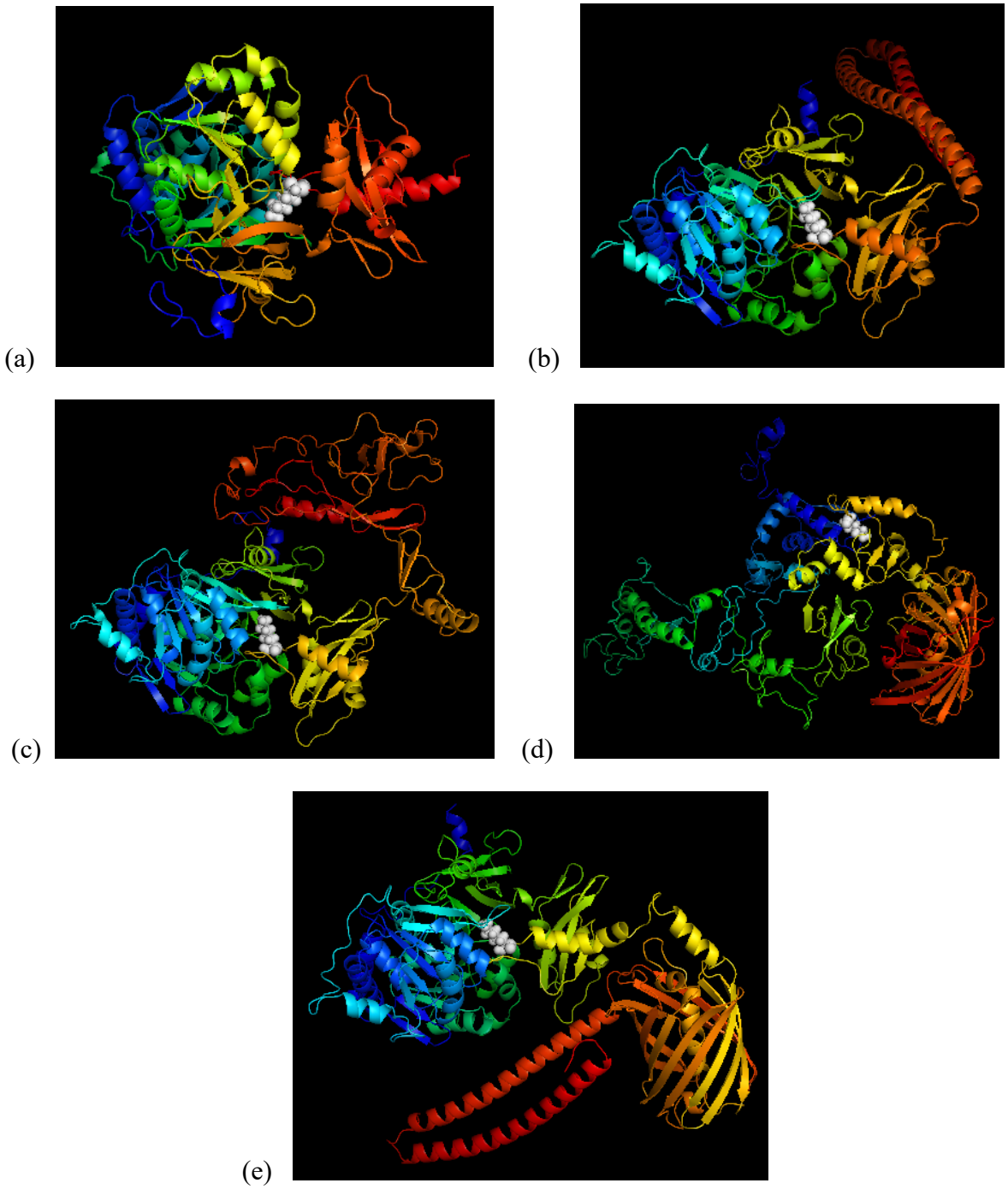


Figure 3.13– Structures of (a) Luc, (b) Luc-Q77, (c) mcRFP unfolded Luc-mcRFP, (d) Luc unfolded Luc-mcRFP, (e) Luc-mcRFP-Q77 displayed by Pymol, predicted by Robetta.[225] The structures are oriented to a view with the best presentation of all domains in each protein.

3.3 Summary

In this Chapter, six recombinant proteins were successfully produced and categorized into two sets. One set was the mcRFP fused proteins, and the other set was Luc fused proteins. The fusion effect on both functional partners were discussed. The results show that both proteins experienced certain conformational changes upon fusion with each other and/or with polyQ, which resulted in changes of the apparent properties. It is found that in general, less fluorescent interference of mcRFP occurs than the bioluminescent decrease of Luc. Fused mcRFP retained a minimum of 40% of fluorescence, whereas the maximum bioluminescent retention is 35%. It is interesting to find that the fusion of the two functional partners, Luc-mcRFP, even illustrates an increase in mcRFP intensity by 25%.

Further analysis of the chromophore of mcRFP and the substrate binding sites of Luc leads to more insights to the structural change upon fusion. In particular, the special absorption spectrum of mcRFP indicated the co-existence of the blue intermediate and the green intermediate during, and even after the maturation of the red chromophore. Fusion to mcRFP altered the favourable maturation pathway of the red chromophore with more blue intermediate left, which shows the role of the extra β-sheet induced by polyQ, or the rigidified structure by Luc which is further supported by the drop in fluorescent lifetime and the arise in QY. While both Luc and polyQ altered the structure of mcRFP, Q77 shows a protection role of Luc from unfolding by the fusion of mcRFP. It is noticed that only mcRFP fusion largely changed the binding affinity of Luc, whereas further fusion of polyQ recovered the binding affinity towards ATP. Furthermore, even though the Luc and mcRFP fusion largely decreased the distance of the donor and the acceptor for BRET, no BRET was observed. This is indicative of a rigidified structure upon fusion of Luc and mcRFP, preventing the correct dipole-dipole coupling.

These apparent property changes could be explained by exploration of the recombinant protein structures, which was achieved through FT-IR, CD, and *de novo* structural prediction. These data calculated that 20% of the β-barrel unfolds in mcRFP when it is fused with Q77; whereas the helical structure increased by 28% upon fusion with Q77. The fusion of Luc with mcRFP is more complicated, and the prediction gives two differently folded structures. The largely changed substrate binding affinity towards luciferase favoured the structure with more Luc

unfolding. It is concluded that the two functional partners have the largest interference with each other when fused together. Such interference, however, may not be detrimental, as not only the fluorescence was increased, but also the bioluminescence was recovered. Certain recovery of Luc folded structure could be explained in the predicted structure, as a spatial effect of the N-terminal polyQ folded just in between the two interfering proteins.

Furthermore, the successful production of different polyQ fused proteins (mcRFP-Q77, Luc-Q77, and Luc-mcRFP-Q77), despite activity loss and partial structural unfolding, allows further study of the polyQ-directed protein self-assembly *in vitro*. Such self-assembly will be compared with the interfacial directed assembly in the next Chapter. The assembly kinetics and further structural change upon protein assembly will be explored.

Chapter 4 Functional Protein Assembly

4.1 Introduction

In the previous chapter, protein fusion effects on functional proteins were discussed. Special attention was paid on how the structural fusion partner, polyQ, altered the structures and the consequent properties of the functional elements. The aim of this chapter is to further investigate the role of the polyQ tag in the process of protein-protein assembly, either through aggregation-based assembly upon incubation under various conditions, or by interfacial-directed assembly at an air-water interface (AWI).

The assembly conditions of polyQ tagged proteins towards fibril aggregated were first investigated. The conditions were further adjusted for the assembly of the functional globular proteins without the polyQ fusion to prove the structural guidance role of the polyQ tag. The structure of the aggregates will be investigated through microscopic and spectroscopic methods; as well as by detecting the aggregate binding towards the thioflavin-T (ThT), which has been used to monitor the amyloid like β-sheet formation in the protein aggregates in the widely-accepted theory.[212]

Apart from the aggregated products, proteins are also known to self-assemble at interfaces due to their amphiphilicity. The assembly of the fusion proteins at an AWI with or without the polyQ tag will be compared in this chapter. The well-developed Langmuir Blodgett technique supports a more detailed investigation of protein assembly kinetics towards the interface as well as the protein conformational changes during the process, comparing to the aggregated assembly. The role of polyQ on different fusion partners during the assembly at AWI will also be determined, based on the different folding structures as was discussed in Chapter 3.

4.2 Results and Discussions

4.2.1 Aggregation-based Assembly of Fusion Proteins

In order to study the role of polyQ during protein assembly, mcRFP with and without the Q77 tag was used as the first model for the investigation of the aggregation process. The aggregation of mcRFP-Q77 was carried out by incubating the protein solution under different pH conditions for a period of time. Aggregates with different shapes appeared after 1-day incubation at room temperature under SEM (Figure 4.1 (a)-(c)). It is previously reported that natural globular proteins may form fibril aggregates at a pH far away from the isoelectric point (IEP) under heat denaturation, comparing to the spherical aggregates near the IEP of the protein.[34] The aggregation of mcRFP-Q77 at different pH agrees with previous findings, even though no heating process is required due to easily aggregated polyQ structure. The calculated IEP value for mcRFP-Q77 is 5.57, and the assembly structure shows a high pH-dependence. A spherical structure was observed at slight acidic condition at pH 5 (Figure 4.1 (b)), when the protein is uncharged. The aggregates appeared as fibril structures up to 2.5 μm in length and 92 nm in width at pH 3 (Figure 4.1 (a)), when the protein molecules are positively charged; whereas rod structures were observed with negatively charged protein molecules (pH 8, Figure 4.1 (c)). The rod-shaped aggregates formed at pH 8 are similar to the literature reported rod shaped protofibril intermediate with around 50 nm in length, and 2.6 nm in diameter during the amyloid fibril formation of Tau protein.[227] To compare the size, the polyQ tagged mcRFP formed the rod shaped intermediates up to 600 nm in length, and 170 nm in width after 1 day of incubation at room temperature, pH 8. As is seen in Figure 4.1 (c) and Appendix A.3 Figure A3.1 (c), these rod aggregates gathered into clusters as they form, which might indicate the tendency of further assembly towards the amorphous aggregate mixture with long fibres after 2-weeks of incubation (Figure 4.1 (d) and Appendix A.3 Figure A3.1 (d-f)).

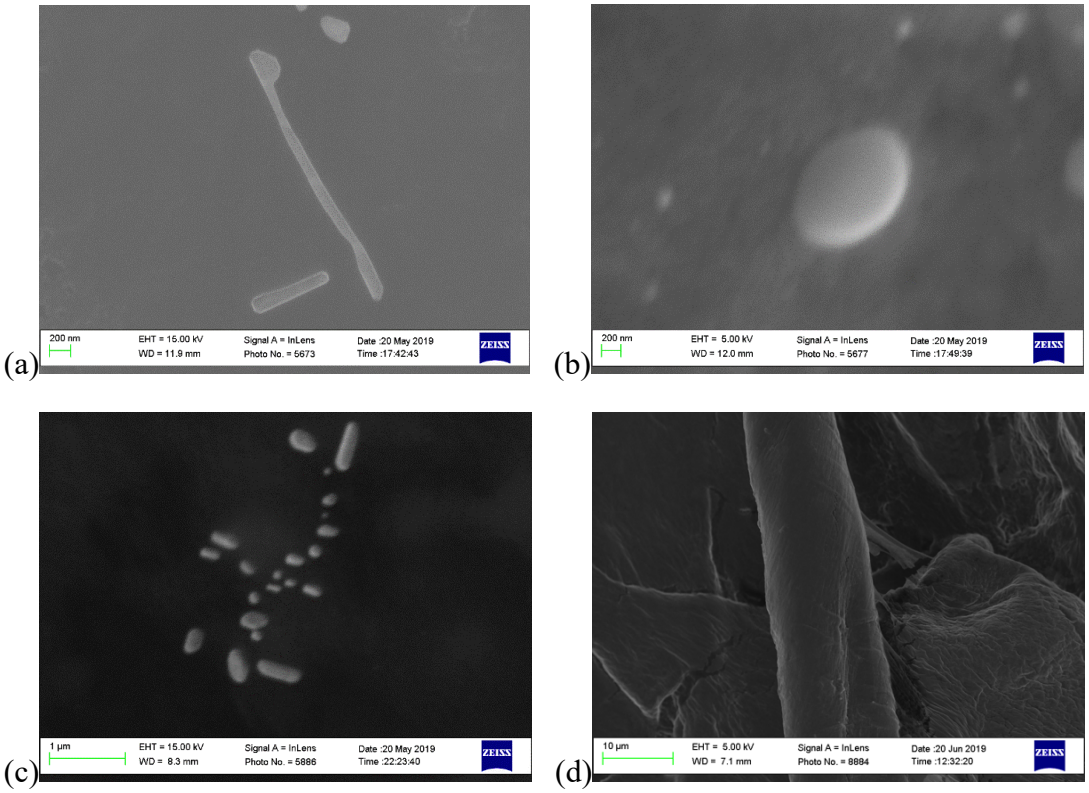


Figure 4.1– SEM images of mcRFP-Q77 self-assembled after 1-day incubation at (a) pH 3; (b) pH 5; (c) pH 8; and (d) after 2 weeks of incubation at pH 8 at 5 µM, room temperature.

Figure 4.2 shows the shape of the aggregates under fluorescent microscope. It could be noticed that after 2 weeks of incubation, nearly no fluorescence is left in the background solution. The aggregated molecules remain fluorescent. Fibrous structures with various lengths up to 400 µm and 18 µm in diameter was observed in the aggregated suspension after 1 week of incubation at room temperature (Figure 4.2 (a)). Further incubation results in large amorphous aggregates with fibrous shape at the edges (Figure 4.2 (b)). Such fibrous shape at the edges matches the long fibres observed under SEM (Figure 4.1 (d) and Appendix A.3 Figure A3.1 (d-f)). These results confirmed a step-by-step formation of protein fibres from the aggregation of Q77 tagged mcRFP. The protein molecules first started to participate into rod-shaped seeds, which gathered together through the intramolecular backbone-sidechain hydrogen-bond[87] and thus long fibres are formed.

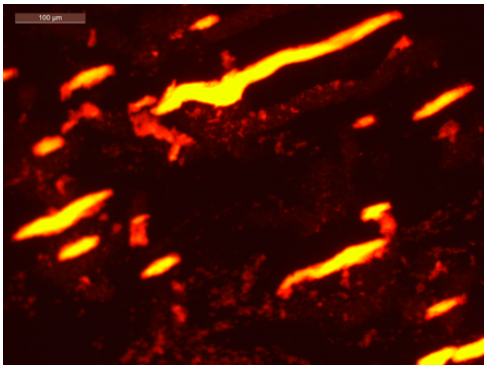
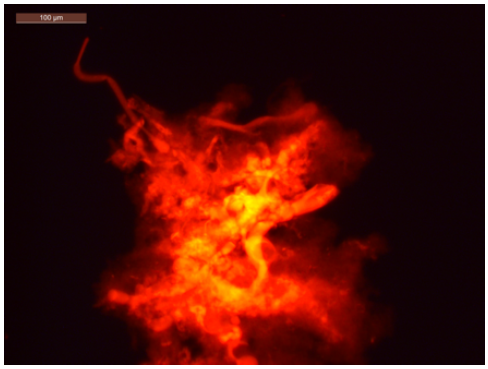
(a)  (b) 

Figure 4.2– Fluorescent images of mcRFP-Q77 aggregates after (a) 1 week and (b) 2 weeks of assembly at room temperature at a concentration of 5 μM, pH 8. (Scale bar: 100 μm)

On the contrary, it is observed under the fluorescent microscope that mcRFP was stable at room temperature even after a long time of incubation. The background solution remained fluorescent regardless of any heating or pH changes (Figure 4.3). The fluorescent microscopic picture shows the fluorescent mcRFP background without any noticeable aggregates after 1 week of incubation at pH 8 (Figure 4.3 (a)). A general decrease in the background fluorescent intensity (darker red colour) was seen at pH 3, showing the denaturing effect at the acidic conditions due to the protonation of the ionogenic group in the red chromophore[228] (Figure 4.3 (b)). In contrast, less denaturing effect was found from the background fluorescence with heating (Figure 4.3 (c)). Few fluorescent aggregates without any obvious fibrous shape were observed by elevating the temperature to 50 ℃ at pH 8 (Figure 4.3 (c)). However, fibrous aggregates up to 668 μm in length, 26 μm in diameter which remained fluorescent were seen after 1 week of incubation at 50 ℃ at pH 3 (Figure 4.3 (d)). The results are slightly different from those discovered in the literature.[34] Even though both conditions (pH 8 and pH 3) are far away from the IEP of mcRFP, fibres only formed at acid condition. It is concluded that both an elevated temperature and an acidic pH away from the physiological pH value is required for the formation of protein fibres from the globular mcRFP. Furthermore, the background solution under this condition also remained fluorescent, which indicates that there are still active mcRFP molecules left in the solution. Comparing to the background solution of mcRFP-Q77 with same starting concentration (5μM), it is seen that nearly no fluorescent molecules are left in mcRFP-Q77 solution. Preliminary conclusion could be drawn here that higher aggregation propensity is seen for the Q77 tagged mcRFP.

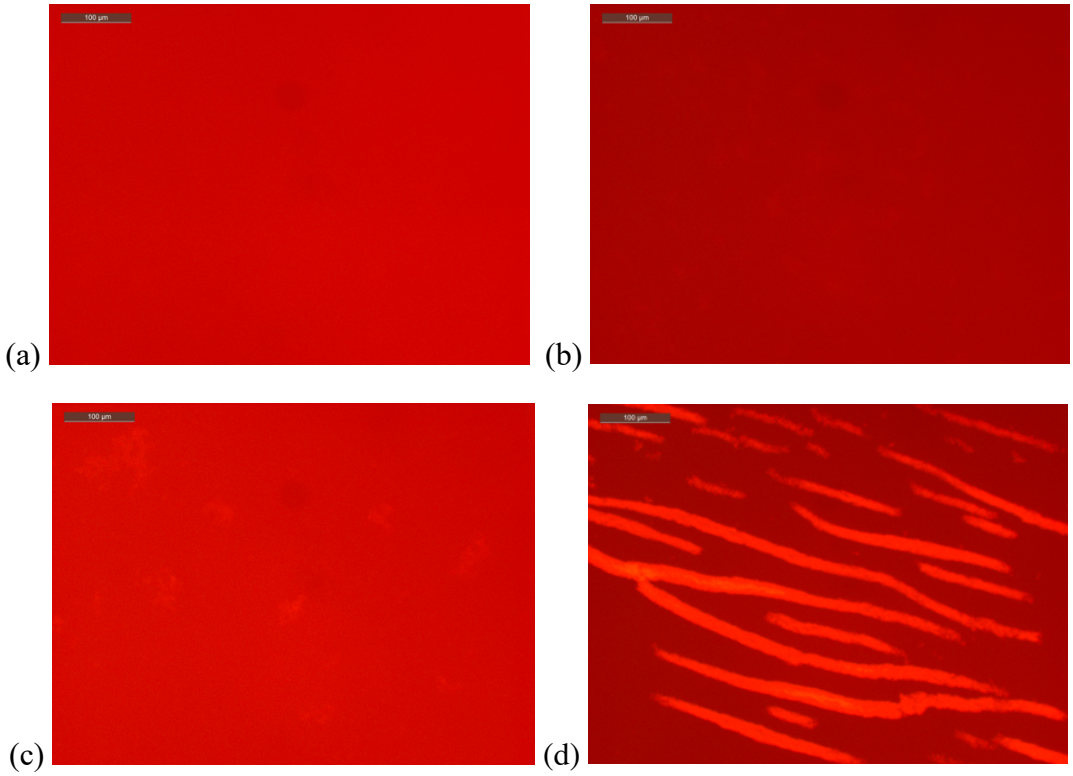


Figure 4.3– Fluorescent microscope images of mcRFP aggregates at a concentration of 5 μM after 1 week of incubation at (a) room temperature, pH 8; (b) room temperature, pH 3; (c) 50 ℃, pH 8; and (d) 50 ℃, pH 3. (Scale bar: 100 μm)

To further study the secondary structure of the aggregates, Thioflavin T (ThT) was used to detect the β-stack structures, which could indicate the formation of the amyloid-like fibril structures in the aggregates. The fluorescent emission from ThT binding with both mcRFP and mcRFP-Q77 was detected after 1 week of incubation, pH 8 (Figure 4.4). The results indicated an increased amount of β-stack structures in mcRFP-Q77 compared with mcRFP, consistent with a structural change of the polyQ region, from the helical structure into β-stack after the protein molecules self-assemble. This result further supports the theory that an intramolecular backbone to sidechain H-bond is formed, contributing to the increased sheet structure.

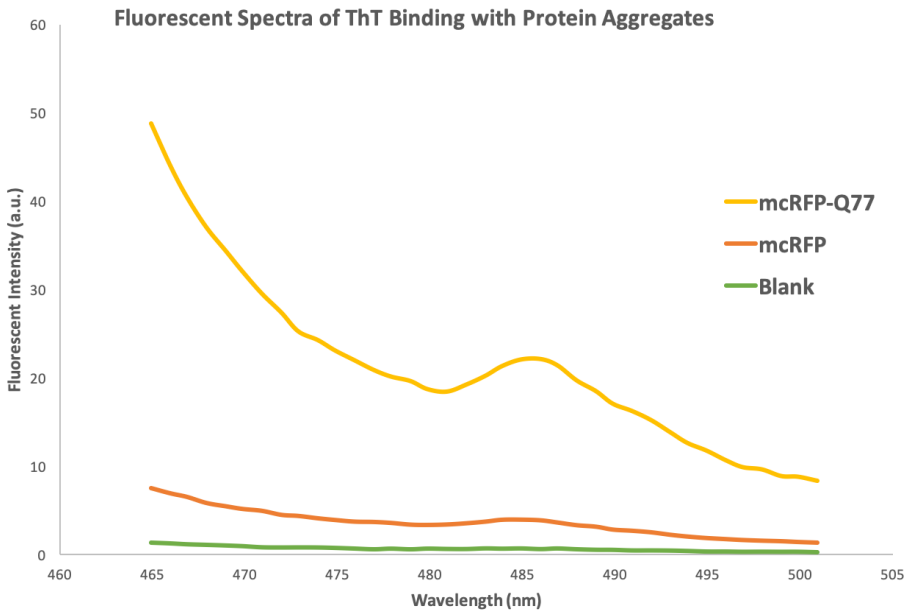


Figure 4.4– Fluorescent emission profile of ThT binding with mcRFP and mcRFP-Q77 aggregates formed from 5 μM protein solution after 1 week of incubation at ambient conditions, pH 8.

The deconvolution fitted FT-IR spectrum of the mcRFP-Q77 aggregates also demonstrated this structural transition (Figure 4.5). The peak for mcRFP-Q77 in solution at 1652 cm^{-1} indicating α-helix structure (Figure 3.9) was red shifted to 1661 cm^{-1} (Figure 4.5), which corresponds to the β-turn structure. On the contrary, the peak was maintained for the β-sheet structure, even though a blue shift from 1632 cm^{-1} to 1624 cm^{-1} was observed. Both positions are in the β-sheet diagnostic region. It is also interesting that the peaks in the assembled structure are broad and overlapped with each other. The recognition of the peaks is only achieved by a deconvolution fit. This is explained by a highly compact structure of the aggregated product, in which the molecular vibration is interacted with the adjacent molecules. The already β-sheet rich mcRFP, however, did not form separated aggregates after months of incubation at room temperature, preventing the parallel comparison. However, the obvious transition from the helical structure towards β-sheet proved the intermolecular hydrogen bond in the mcRFP-Q77 aggregates.

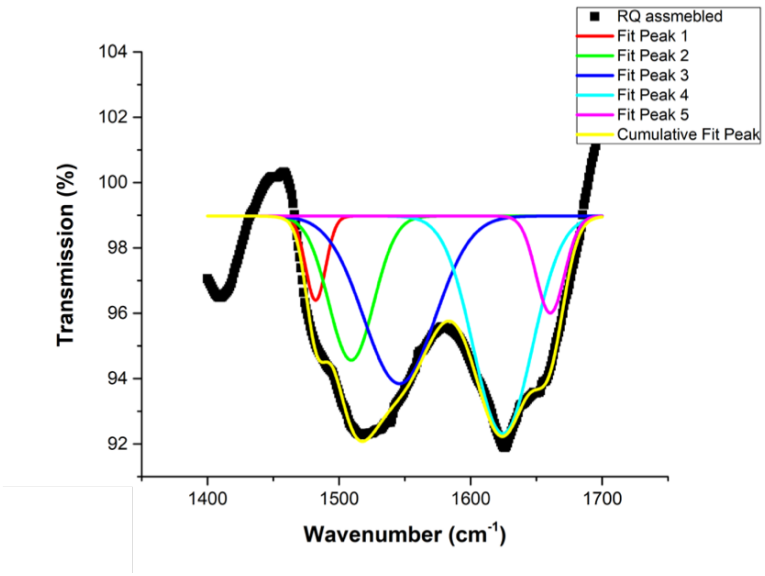


Figure 4.5– FT-IR spectrum of mcRFP-Q77 aggregates formed after 1 week of incubation at 5 μM, ambient conditions, pH 8. Deconvolution fit was performed with Origin.

The secondary structures of these aggregates were further confirmed by the ThT fluorescent spectrum (Figure 4.5). However, ThT has been shown to bind poorly at pH<3.[212] Therefore, only the structure formed at pH 8 was confirmed using this binding fluorescent method. This binding fluorescence at pH 8 was also compared at different incubation times, and the emission peak shows an increase in the fluorescent intensity with longer incubation, indicating the increase in the β-sheet structure over the time. These sheet contents, however, grow faster within the first two days, as the increase of the intensity was smaller after 2 days within a week (Figure 4.6). The results also match the microscopic images from short rod aggregates (Figure 4.1 (c)) to long fibres (Figure 4.2).

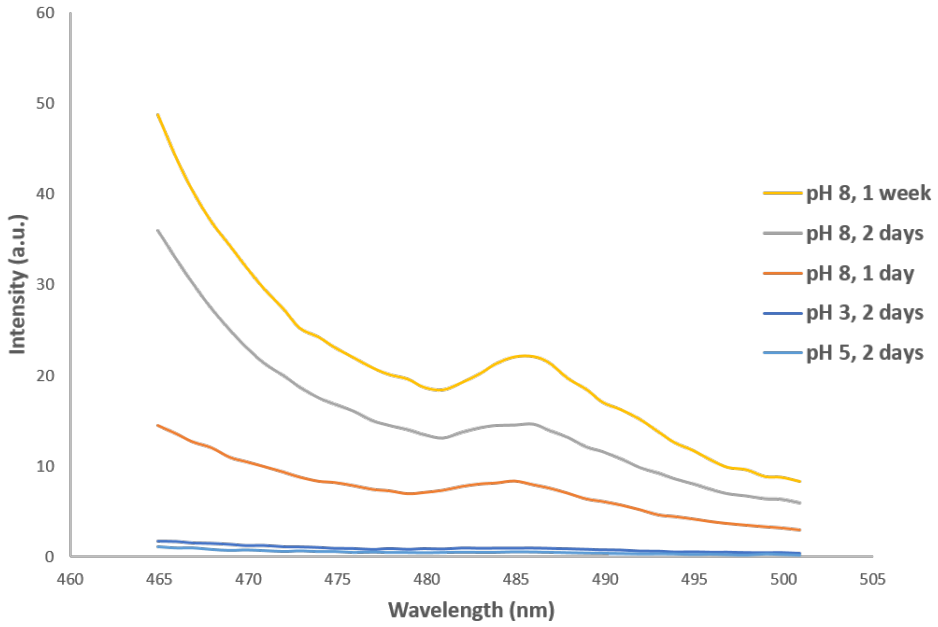


Figure 4.6– Fluorescent emission profile of ThT binding with mcRFP-Q77 aggregates from 5 μM protein solution at different pH for different time.

With Luc constructs, it is interesting that a different assembly behaviour was observed from the ThT binding fluorescent spectra (Figure 4.7). Similar to the mcRFP-Q77, the β-stack contents in Luc-Q77 (LQ) increase with increasing incubation time, detected by ThT binding assay. However, it is interesting to find that Luc-mcRFP aggregates with a high β-stack content after 1 month of incubation. On the contrary, the polyQ tagged protein, LRQ, did not show any β-stack development after 2 months of assembly. This might be explained by the different structural change as discussed in Chapter 3 (Figure 3.13). The Luc-mcRFP experienced the largest degree of structural change, which might lead to the aggregation upon long time incubation. The polyQ in the LRQ, however, was spatially surrounded by two large functional proteins, which is potentially preventing the intramolecular sidechain to backbone interaction among polyQ domains. Therefore, the LRQ protein shows structural stability with incubation. Slight increase in β-stack content was also observed for Luc with 2 months of incubation. This indicates an unstable structure of the Luc protein on itself, which was stabilised by the mcRFP-Q77 fusion.

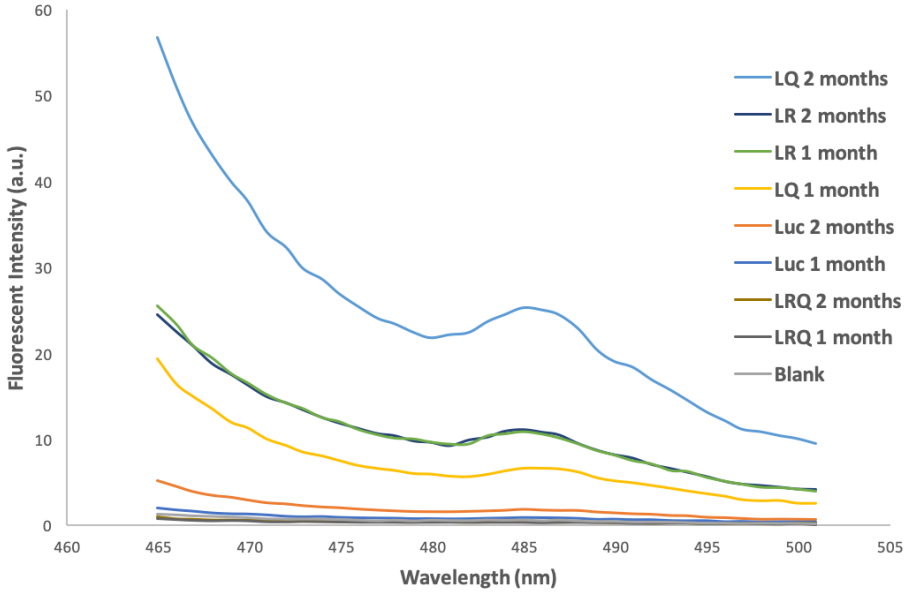


Figure 4.7– Fluorescent emission profile of ThT binding with different Luc constructs incubated for different time at pH 8, room temperature.

4.2.2 Fusion Protein Assembly at an Air-Water Interface

While aggregation of protein molecules reported in the literature usually require harsh conditions such as chemical denaturants, high temperature or low pH, the assembly of the recombinant proteins with different fusion could be achieved under more ambient conditions as reported in the previous section. These assembled aggregates, however, were obtained after a long incubation time, and the structure is non-uniform and hard to control. Directed assembly such as electrospinning under an applied high voltage, or the most recent developed printing techniques could confine the assembly product to certain desired size and shape. These methods largely reduce the activity of proteins due to the high voltages and the shear stress applied to the proteins; and are therefore mostly studied for their use in structural protein scaffolds. The hydrophobicity directed protein assembly at an air-water interface (AWI), in these cases, is under milder conditions than the electrospinning and printing techniques and more controllable comparing to the aggregation method. In this chapter, the assembly at AWI will be used on recombinant proteins with or without a polyglutamine tag under mild conditions.

4.2.2.1 Effect of polyQ on mcRFP at the AWI

As soon as protein solution was dropped in water, the assembly of protein molecules towards the interface began. However, it may take some time for the surface pressure to start to increase. Figure 4.8 compares the assembly profile of mcRFP and mcRFP-Q77 at different initial protein concentrations and at different total molecules of protein. These curves could also be represented by plotting the log (d Π /dt) over Π (Appendix A.4 Figure A4.3), as was described in Figure 1.11 (b). This log (d Π /dt) over Π curve demonstrate the rate of assembly at each surface pressure. It could be noticed in Figure 4.8 that there might be more than one time point at a certain surface pressure, and thus more than one assembly rate exists as is shown in the Appendix A.4 Figure A4.3. The multiple log (d Π /dt) values at certain Π values reflect the unstable surface with fluctuating surface pressure. An important parameter in this log (d Π /dt) over Π curve was the initial constant fast assembly rate, which is listed in Table 4.1. In general, the assembly of mcRFP-Q77 was faster, which resulted in a higher final surface pressure than mcRFP at same concentration and same number molecules of protein. In each case, it is noticed that the final surface pressure is depended on both these parameters.

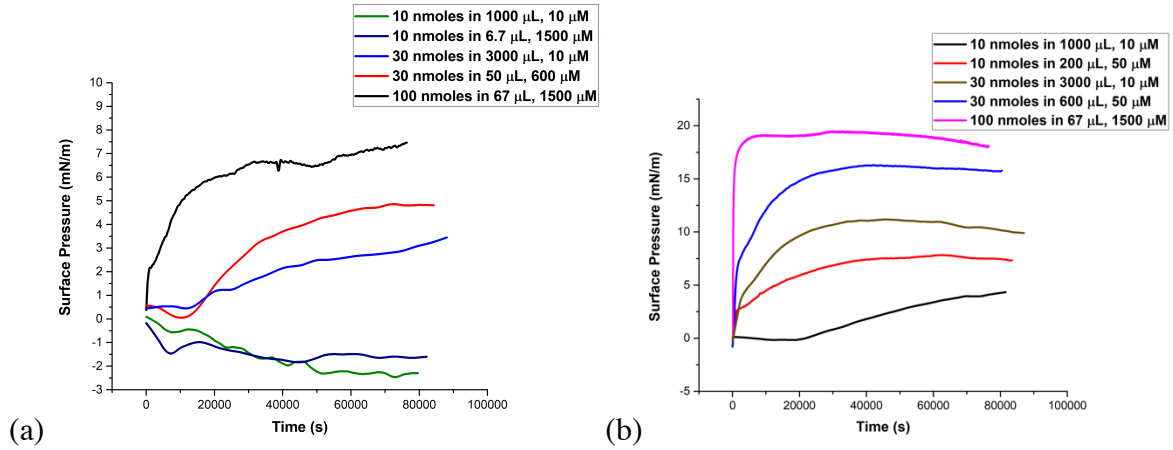


Figure 4.8– The assembly profile (Π-t plot) at the interface of air and pure water of (a) mcRFP and (b) mcRFP-Q77 in different molecular quantity at different initial concentration before spread on the interface.

Table 4.1 – Initial constant assembly rate of mcRFP and mcRFP-Q77 at different molecular quantity and different initial concentration.

	Initial constant assembly rate (Log(mN/m·s))	
	mcRFP	mcRFP-Q77
10 nmoles in 1000 μL, 10 μM		-4
10 nmoles in 200 μL, 50 μM		-2.5
30 nmoles in 3000 μL, 10 μM	-4	-3
30 nmoles in 600 μL, 50 μM		-2.4
30 nmoles in 50 μL, 600 μM	-3.6	
100 nmoles in 67 μL, 1500 μM	-2.5	-1.5

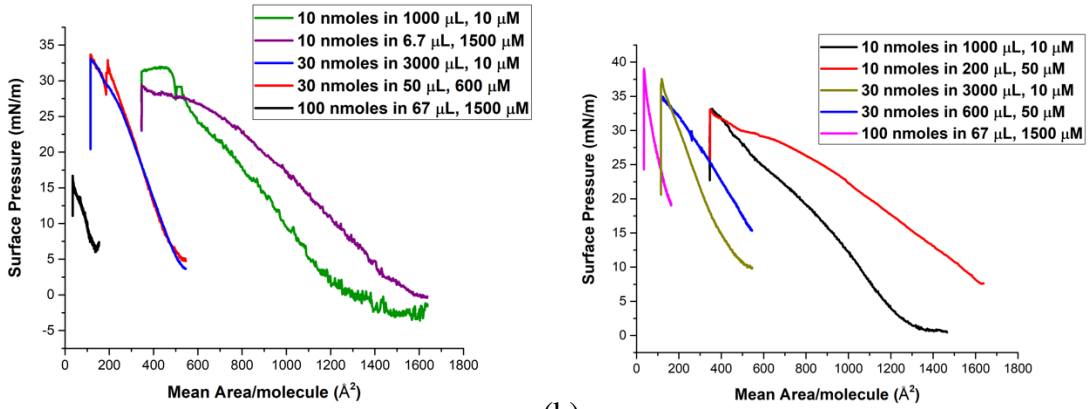
The assembly of mcRFP at AWI in Figure 4.8 (a) shows an unstable surface pressure change. The surface pressure did not increase when only 10 nmoles of protein molecules were added to the interface, regardless of the concentration. In this case, the protein adsorbed on the surface as gaseous molecules without intermolecular interaction between each other.[229] Proteins adsorbed at the surface as isolated molecules when only a very small amount of proteins is spread. Without the interaction among the adsorbed molecules, the Wihelmy plate cannot detect the pressure change of such a low molecular density surface. When the total mcRFP molecules increased to 30 nmoles, the surface pressure started to rise after a delay of 15,000 seconds. It is noticed that the delay time is roughly the same at initial concentrations of both 10 μM and 600 μM. In theory, the assembly process for these two conditions should be different. At a higher initial concentration, when the protein solution was added in the water, protein molecules rapidly assembled to the surface at $10^{-3.6}$ mN/m·s (Table 4.1). These molecules diffused across the whole surface. In comparison, when protein solution was added in the water at a lower initial concentration (larger volume), protein molecules gradually adsorb on the surface at 10^{-4} mN/m·s (Table 4.1) as the gaseous phase. Protein molecules have the

least tendency to interact with each other in both cases until there is no more "free area" for another gaseous molecule. Further assembly resulted in protein molecular interactions and thus the increase in surface pressure. Even though both processes before molecular interactions take the same time, the molecular unfolding status is already different. Therefore, further assembly follow different kinetics, which resulted in a higher final surface pressure for the higher initial concentration. With increasing mcRFP molecules to 100 nmoles at 1500 μM initial concentration, the protein molecules assemble and diffuse so rapid at $10^{-2.5}$ mN/m·s (Table 4.1) that the delay period could be neglected.

For the case of mcRFP-Q77 (Figure 4.8 (b)), the curve of surface pressure increase did not experience much fluctuation. The addition of 10 nmoles of protein molecules resulted in an increase in surface pressure. Comparing with mcRFP, the results indicate a higher tendency of mcRFP-Q77 molecules to assemble towards the interface, which might be due to the stronger interactions among mcRFP-Q77 molecules. It is therefore deduced that the mcRFP-Q77 molecules tend to form intramolecular H-bond to direct the whole molecule towards the surface and to rapidly interact with each other on the surface. At lower initial concentration, the molecules could still assemble slowly at 10^{-4} mN/m·s (Table 4.1) as gaseous phase before they started to interact with each other, given that the total number of protein molecules is low enough. At higher concentration, however, the long delaying time no longer exists. Molecules assembled fast at $10^{-2.5}$ mN/m·s (Table 4.1) to the surface and rapidly interact. Again, the increased initial concentration with the same number of total molecules in the system resulted in a higher final surface pressure, which means different unfolding status of the protein molecules at the surface. More obviously, larger number of total molecules resulted in the faster initial assembly rate, and the higher final surface pressure due to the more crowded surface and thus stronger molecular interaction. This is explained as the intramolecular H-bond formed with the Q77 backbone and side group expels the surrounding water molecules, which forces the molecule towards the interface.

After the surface pressure was stable at the interface, the surface was compressed for the production of a condensed film, whose characteristics and the consequent fibre formation will be discussed in the next Chapter. During the compression of the surface filled by interacted protein molecules, further increase in surface pressure was noticed by recording the Π-A

isotherm (Figure 4.9). The final drop at the smallest area in each curve indicates the process of fibre pulling. The Π-A isotherm is widely used to indicate the status of the film on the interface, which demonstrates the molecular gas, liquid (including liquid expanded and liquid condensed), and solid phases.



(a) (b)

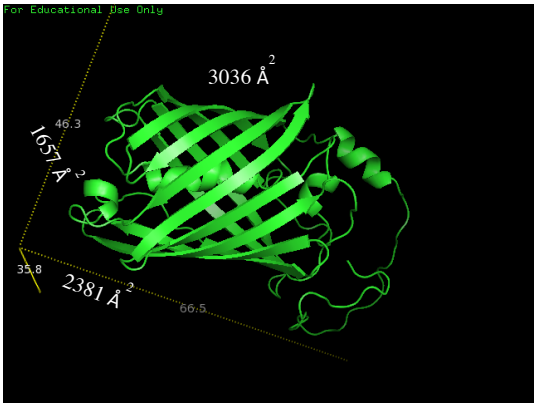
Figure 4.9– The compression isotherm (Π -A plot, A is the mean area each molecule occupies in Å²) at the AWI of (a) mcRFP and (b) mcRFP-Q77 at different molecular quantity and different initial concentration, under a compression speed of 100 cm²/min with two barriers.

Limiting areas for each phase transition was calculated using the method described in Figure 1.12. The calculated numbers for all protein constructs were listed in Appendix A.4 Table A4.1. As is discussed earlier, there was no surface pressure increase when 10 nmoles of mcRFP was added, regardless of the concentration. However, a difference is observed in this Π-A plot (Figure 4.9 (a)), which indicates the gaseous molecules on the surface was compressed to interact with each other. A full monolayer isotherm was observed with 10 nmoles of mcRFP at 10 μM initial concentration (green line in Figure 4.9 (a)). The limiting area for the transition between the gas phase (G) to the liquid phase (L) was calculated as 1235 Å²/molecule, whereas the transition area from liquid expanded phase (LE) to the liquid condensed phase (LC) was 784 Å²/molecule. The solid phase (S) came after a little spike in the isotherm, starting from an area of 688 Å²/molecule, followed by the collapse of the monolayer at 466 Å²/molecule and 32 mN/m. With neither increased amount of protein, higher initial concentration, nor the polyQ tagged construct, full monolayer isotherms were observed. With 10 nmoles of mcRFP at 1500 μM, the limiting areas are larger than that for the low concentration. The transition area from

G to L was increased to 1527 Å^2/molecule, and that for LE to LC was 807 Å^2/molecule. This isotherm did not experience any LC to S phase transition. These results indicated the different orientations of protein molecules at the AWI from different starting concentration, resulting in different interactions among these molecules. With increased amount of protein molecules to 30 nmoles, compression started after the molecules interacted with each other and stabilised at the interface. The further increased surface pressure indicated stronger intermolecular forces and the film started to form clumped structures. The clumped structure could also explain the disappearance of the monolayer collapse, due to the stronger intermolecular forces formed by these aggregates. A largely decreased surface pressure after compression was observed for 100 nmoles mcRFP at 1500 μM, even though the starting pressure was the highest (Figure 4.14 (a)). This could be explained as an over-crowded interface, from which proteins started to desorb back to the bulk during compression.

The larger phase transition areas for mcRFP-Q77 could be resulted from a larger molecular size Appendix A.4 Table A4.1. It was previously reported that α-helical peptides could assembled at an air-water interface both in perpendicular and parallel directions. Compression of the parallel helices resulted in a transition towards β-sheet structure. However, the assembly orientation for a protein is more complicated. From the difference at different molecular quantities and different concentrations, it is also deduced that the assemble orientation is changed from case to case. Again, the disappearances of monolayer collapse for mcRFP-Q77 at all conditions show larger intermolecular interactions at the interface, which might also be resulted from the intramolecular H-bond by polyQ with the "clump" effect. Different from mcRFP, the compression of 100 nmoles mcRFP-Q77 at 1500 μM resulted in a highest surface pressure. This might indicate the transition of the single helical Q77 towards intermolecularly interacted β-sheet structure after surface compression.

Figure 4.10 shows the theoretical dimensions of mcRFP and mcRFP-Q77. The calculated limiting areas in all cases are smaller than every cross-sectional area of the corresponding protein. This indicates molecular rearrangement occurs with conformational changes for all of the conditions.

(a) 

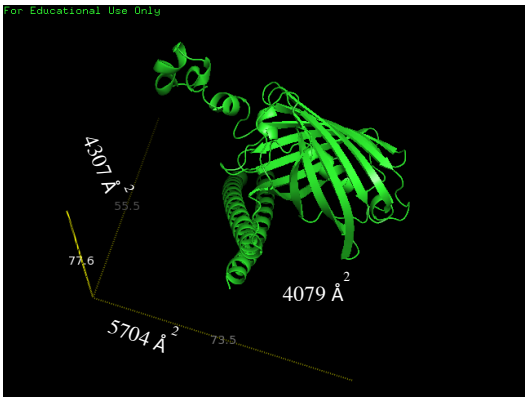
(b) 

Figure 4.10– Theoretical dimensions of (a) mcRFP and (b) mcRFP-Q77 labelled with cross-sectional areas (in Å^2) for all three directions of viewing. (calculated and displayed by Pymol)

4.2.2.2 Effect of polyQ on Luc at the AWI

In comparison, 10 nmoles of Luc on its own assembled fast at the AWI with the final surface pressure being able to reach 18 mN/m (Figure 4.11 (a), black and red lines). This further proves the fact that Luc is more facile to structural conformational changes. Therefore, the Luc molecules could easily rearrange themselves towards the interface. The assembly profile of Luc, which is different from mcRFP, follows several stages: the initial fast assembly period, the first short stabilising period, the second assembly period with reducing rate, and the final stabilising period. It is noticed that when the total amount of luciferase molecules increased to 30 nmoles, a decay in surface pressure occurred from 20 mN/m down to 15 mN/m (Figure 4.11 (a), blue line). This decay was followed by a further increase after a transition point at around 65,000 seconds. Such a decay might indicate either a fast molecular rearrangement on the interface, which could not be compensated immediately by the further molecular adsorption towards the surface; or a desorption from the surface due to the largely crowded interface, which is accompanied by molecular rearrangement until a point where the remaining molecules adopted a certain alignment for further molecular adsorption. The final increase, on the other hand, is resulted from the increase in the molecular interactions. Similar decaying profile is also observed with Luc-mcRFP construct without the further increase (Figure 4.11 (b)). It is noticed that for Luc-mcRFP, an increase in either protein amount or initial protein

concentration results in a lower final surface pressure after assembly. The results show that even though Luc-mcRFP assembles resulted in higher final surface pressure than mcRFP with the same number of protein molecules, the film was not stable with a decay in surface pressure, which might indicate some desorption or molecular rearrangement at the interface.

The behaviour of Luc-Q77 and Luc-mcRFP-Q77 are different. With 10 nmoles of Luc-Q77 at 10 μM, the molecules formed a "gaseous" phase at the interface. No increase in the surface pressure was observed, as was shown for mcRFP at the same molecular amount and initial concentration. A larger molar number at higher initial concentration results in rapid assembly until 10 mN/m (Figure 4.11 (a)) before the stabilisation period. The overall profile for 30 nmoles Luc-Q77 is similar to mcRFP-Q77. Therefore, Q77 plays a role in directing the assembly of Luc. In comparison, 10 nmoles of Luc-mcRFP-Q77 at 10 μM initial concentration could assemble at the interface. Both the assembly rate and the final surface pressure are smaller than Luc-mcRFP at the same conditions (Figure 4.11 (b) red line comparing to the black line).

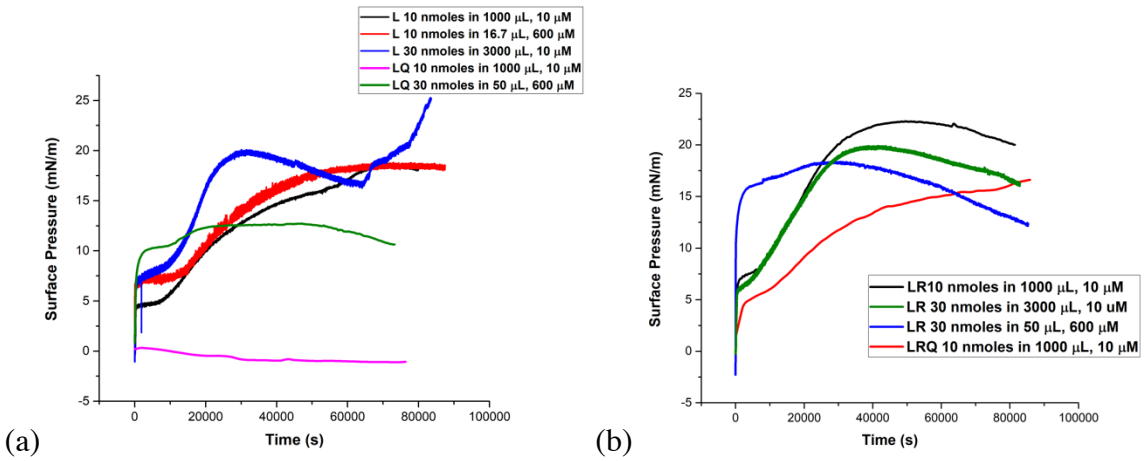


Figure 4.11– The assembly profile (Π-t plot) at the AWI of (a) Luc and Luc-Q77, (b) Luc-mcRFP and Luc-mcRFP-Q77 in different molecular quantity at different initial concentration.

The detailed assembly rate at each surface pressure could be obtained from the kinetic profiles Appendix A.5 Figure A5.5. Again, the initial rate of assembly is constant at each condition, which could be compared in parallel (Table 4.2). For Luc and Luc-mcRFP without polyQ tag, it is shown that the initial rate of assembly is independent of the total number of protein

molecules, but increased with higher protein concentration, due to the faster diffusion. The assembly rates for the two proteins at the same concentration are the same but are different from that of mcRFP. This may suggest the dominate role of Luc during the assembly at AWI. Both Luc-Q77 and Luc-mcRFP-Q77 assembled slower than its non-polyQ tagged counter partner. These results show that the effect of polyQ on the assembly of Luc and mcRFP are controversial. Therefore, it is concluded that the assembly of polyQ depends on its fusion partner, as is reported in literature.[73]

Table 4.2 – Initial constant assembly rate of Luc constructs at different molecular quantity and different initial concentration.

	Initial constant assembly rate (Log(mN/m·s))			
	L	LR	LQ	LRQ
10 nmoles, 10 μM	-2	-2		-3
10 nmoles, 600 μM	-0.8			
30 nmoles, 10 μM	-2	-2		
30 nmoles, 600 μM		-0.8	-1.8	

Figure 4.12 shows the phase transitions during compression for each Luc fusion constructs at different molar amounts and different concentrations. The calculated limiting area of each phase transition are listed in Appendix A.5 Table A5.2. 10 nmoles Luc at either 10 μM or 600 μM initial concentrations resulted in the same final surface pressure after assembly (Figure 4.11 (a)). However, a liquid to solid transition with a collapse of the interfacial film was only observed at the lower concentration (Figure 4.12 (a)). In comparison, the compression of the 10 nmoles of Luc at the higher concentration stopped at the liquid condensed phase. This result is the same with the one obtained with mcRFP, where 10 nmoles of protein at lower concentration resulted in a collapse of the film after the solid states. As was discussed earlier, these are resulted from a different molecular assembly and arrangement at the interface at different concentration. At lower concentration, molecules assembled relatively slower than

those at higher initial concentration. The results show that molecular interactions from these slowly adsorbed molecules are stronger. A large spike was observed with elevated amount of Luc, which is possibly indicating molecular desorption from an overcrowded surface. The remaining molecules quickly rearrange for further assembly of more molecules and result in the recovery of the surface pressure to form a solid-state film.

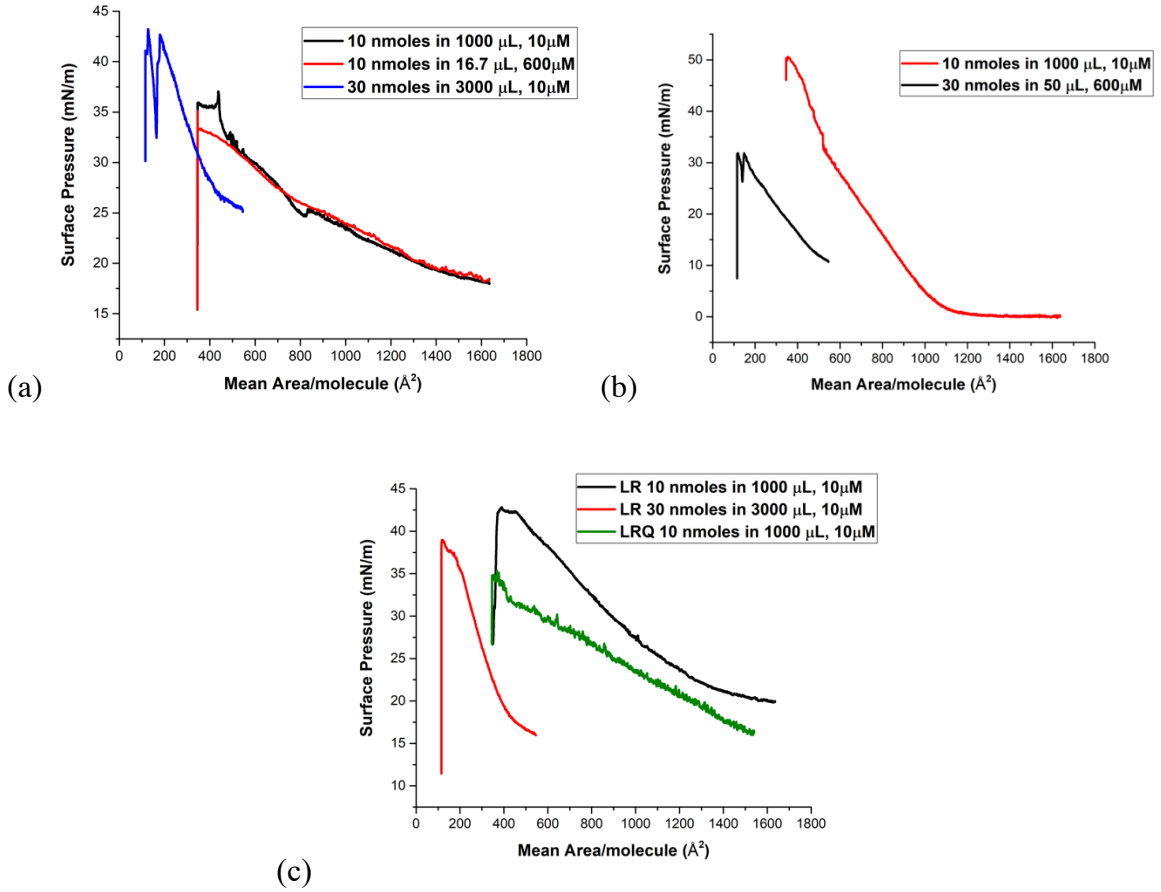


Figure 4.12– The compression isotherm (Π -A plot, A is the mean area each molecule occupies in Å^2) at the AWI of (a) Luc, (b) Luc-Q77, (c) Luc-mcRFP and Luc-mcRFP-Q77 at different molecular quantity and different initial concentration, under a compression speed of 100 cm^2/min with two barriers.

A full isotherm was also observed for the 10 nmoles of Luc-Q77 at 10 μM (Figure 4.12 (b)). The transition from gas phase to liquid expanded phase happened at 1082 Å^2/molecule. Its transition towards the liquid condensed phase matches the behaviour of the reported β-sheet peptides, which forms a stable LC phase until a high pressure near 50 mN/m.[230] Therefore, it is

highly possible that a polyQ tagged Luc also formed β-sheet structure at the air-water interface as mcRFP did, even though the influence of polyQ towards the assembly rate of Luc and mcRFP are opposite. More protein molecules (30 nmoles at 600 μM) resulted in a spike again during compression. The spike, together with the much lower solid-state transition area, indicates a crowded film. The lowered surface pressure, however, might be indicative of less β-sheet structure comparing to the smaller molar amount (Figure 4.12 (b) black line comparing to the red line).

The compression of 10 nmoles Luc-mcRFP at 10 μM resulted in the film collapse again (Figure 4.12 (c) black line). Larger quantity of proteins did not form an over-crowded film as Luc did (Figure 4.12 (c) red line). However, the much smaller footprint at transition (189 Å^2/molecule) indicates a more compressed structure, showing the higher compressibility of Luc-mcRFP comparing to Luc. The solid state was also noticed for Luc-mcRFP-Q77, without an obvious collapse (Figure 4.12 (c) green line). The LE to LC transition happens at 988 Å^2/molecule, which is also much smaller than that of Luc-mcRFP (1751 Å^2/molecule). These results indicated the even higher compressibility of Luc-mcRFP-Q77 than Luc-mcRFP. Similar to the pair of mcRFP, the transition areas are much smaller than the calculated areas of the 3D protein structure for each construct (Figure 4.13), indicating the change in the molecular structures.

In general, the missing of the film collapse of Q77 tagged molecules (mcRFP-Q77, Luc-Q77, and Luc-mcRFP-Q77) provides a possible insight of the intermolecular backbone-side chain H-bond formation by polyQ. It was also noticed that proteins at different quantity and different concentration assembled, arranged and folded differently. Protein fibres, produced from these films, will be collected from 30 nmoles of protein at 10 μM concentration. The characters will be discussed in the next chapter.

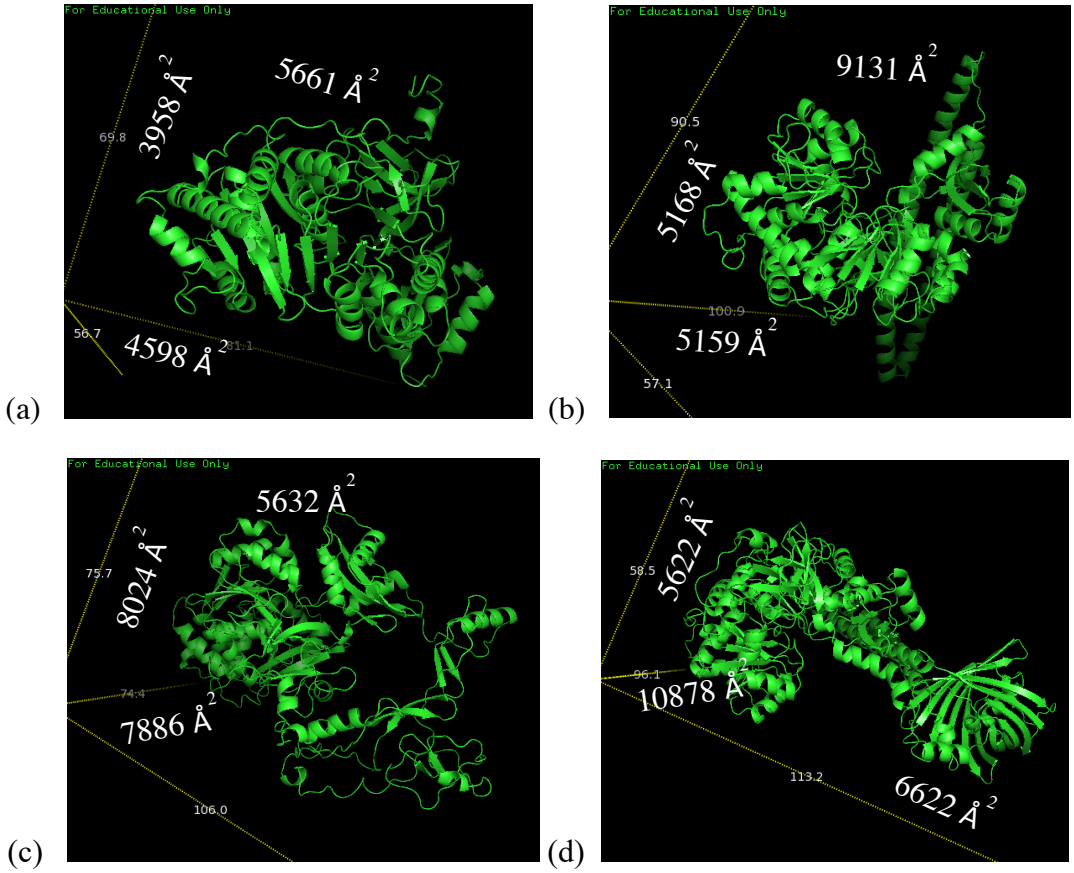


Figure 4.13– Calculated surface areas on each side views of the 3D structures for (a) Luc, (b) Luc-Q77, (c) Luc-mcRFP, and (d) Luc-mcRFP-Q77.

4.3 Summary

In this Chapter, two assembly methods were studied for the six protein constructs. The self-assembly of the globular mcRFP without the polyQ tag requires heat denaturation before assembly to protein aggregates. In contrast, the polyQ assembly occurs in ambient condition with long time incubation. At a pH close to pI, mcRFP-Q77 assembles into a spherical shape indicated by both SEM and ThT binding fluorescent spectroscopy; whereas at pH far away from pI, either rod shaped, or fibre shaped aggregates were observed. The assembly of Luc constructs were also studied by ThT binding. A higher propensity of fibril assembly was observed with Luc-Q77, as well as Luc-mcRFP, compared with the polyQ tagged Luc-mcRFP-Q77. This unusual behaviour was explained by the modelled structure in Chapter 3, that the polyQ tag was spatially protected by the two functional partners, which could prevent the assembly of the Q77 tagged molecules.

The interfacial assembly, however, requires less harsh condition with shorter incubation time. The Wihelmy plate method monitors the process of protein assembly towards the interface. From the readings of the surface pressure change and the kinetic plot, it is possible to get the in-depth information of assembly, such as the molecular adsorption and desorption, the structural change, and the film arrangements. With the six different recombinant proteins, different behaviours were observed during assembly and film compression. The results showed the polyQ tag facilitates the assembly of mcRFP, indicated by both the Π-t plot and its logarithm derivatives (the kinetic plot). The assembly profile for Luc fused constructs are largely different. Luc on its own demonstrated a fast assembly at the AWI. The initial constant assembly rate was dependent on the initial protein concentration, which is explained by a faster diffusion occurred with higher initial concentration. Consequently, the following molecular rearrangement was affected and behaved differently for different protein amount and protein concentration, which was indicated by the compression profile. In general, higher protein amount with lower initial concentration resulted in a film formed by molecular “clumps”. The intermolecular bonds within such a film are stronger, and thus eliminating the appearance of the collapse of a monolayer. However, even though at small protein amount and high concentrations, no collapse appeared in the compression profile of the polyQ tagged constructs. This indicates the intermolecular H-bond formation by polyQ to form aggregates and to stabilise the compressed film from collapse. In conclusion, Q77 facilitates the assembly of

mcRFP; and despite a lower assembly rate was obtained with Luc-Q77 and Luc-mcRFP-Q77, the film was stabilised by the strong H-bond formed.

Comparing the self-aggregation and the interfacial assembly, the milder conditions provide the interfacial assembly with the advantage to further investigate the properties of assembly product, which is to be discussed in the next chapter.

Chapter 5 Functional Protein Fibres

5.1 Introduction

Even though the behaviour of proteins at the air-water interface is still under investigation, the AWI remains a unique tool for guiding proteins as small building blocks into larger self-assembled structures.[231] Most research work focuses on the study of a monolayer protein film, and its deposition on various substrates. Studies of protein structural change with such interfacial assembly were also carried out extensively in the past, mostly with the globular proteins such as lysozyme, bovine serum albumin (BSA), lactoglobulin, and myoglobulin.[202,203] The combination of the functionality and such large structure production, however, was limited. The production of functional silk-like protein fibres has been achieved previously at the air-water interface with Ultrabithorax as the fusion partner, a *Drosophila* transcription factor that self-assembles into materials *in vitro*. This study did not compare the interfacial behaviours and fibril characteristics of globular proteins with and without a self-assembled tag; and the details of protein structural change after fibre formation was not focused.[191,192]

In this study, self-supportive protein fibres with various functionalities were fabricated in ambient conditions via the thin-film route. The thin film was first generated from the self-assembly of different recombinant proteins on an air-water interface (AWI). Such film could then be lifted from the interface into long fibrous materials. In this chapter, mcRFP and mcRFP-Q77 will be taken as modelled protein pairs for the study of the physical properties of the film and fibre, followed by the investigation of their fluorescence recovery in the fibre, compared with the bioluminescent activity in Luc contained fibres. The development of different activity detection for these unique fibrous materials have been developed and evaluated for different types of fibres. The protein structural change will also be studied in these fibrous products. The

contribution of both the structural change and the assembly behaviour towards the protein activity change in the fibre will also be discussed.

5.2 Results and Discussions

5.2.1 Protein Film Formation

Proteins were dropped in water for assembly at the AWI as described in the previous chapter. The proteins formed a layer of gelatinous film at the interface after a period of time, which could be lifted with a b-shaped wire (Figure 2.5). The film, stuck inside the round area of the b-shaped wire, could stay for several minutes depending on the protein type and the surface pressure when it was lifted, before collapsing. Proteins were left assemble on the interface until the surface pressure reached 7 mN/m before the film was lifted in the b-shaped wire at 7 mN/m. A smooth surface of mcRFP film with a cracked layered structure near the edge of the film was found under SEM (Figure 5.2 (a)). No fluorescence was found with this film. However, when the same mcRFP film was compressed by the barriers until 30 mN/m, “wrinkles” appeared (Figure 5.2 (b)). These wrinkles, however, did not align in one direction. This is explained as the force applied to the interfacial film changed its direction when the barrier closes, which was due to the round shape of the LB trough. These wrinkles in random direction were also shown to retain the fluorescence under the fluorescent microscope (Figure 5.1). Therefore, it is concluded that the compression process concentrated the film and improved the properties of the film showing fluorescent fibrous structures.

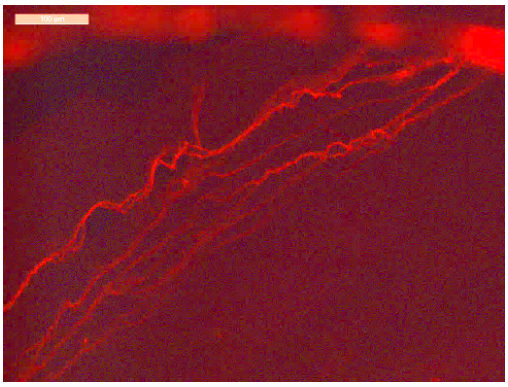
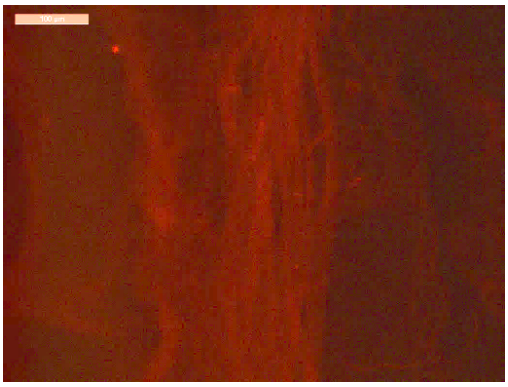
(a)  (b) 

Figure 5.1– Fluorescent microscope images of central mcRFP protein film in the round area of the b-shaped wire at 30 mN/m after compression. (Brightness of the images were adjusted for better visualisation, scale bars: 100 µm)

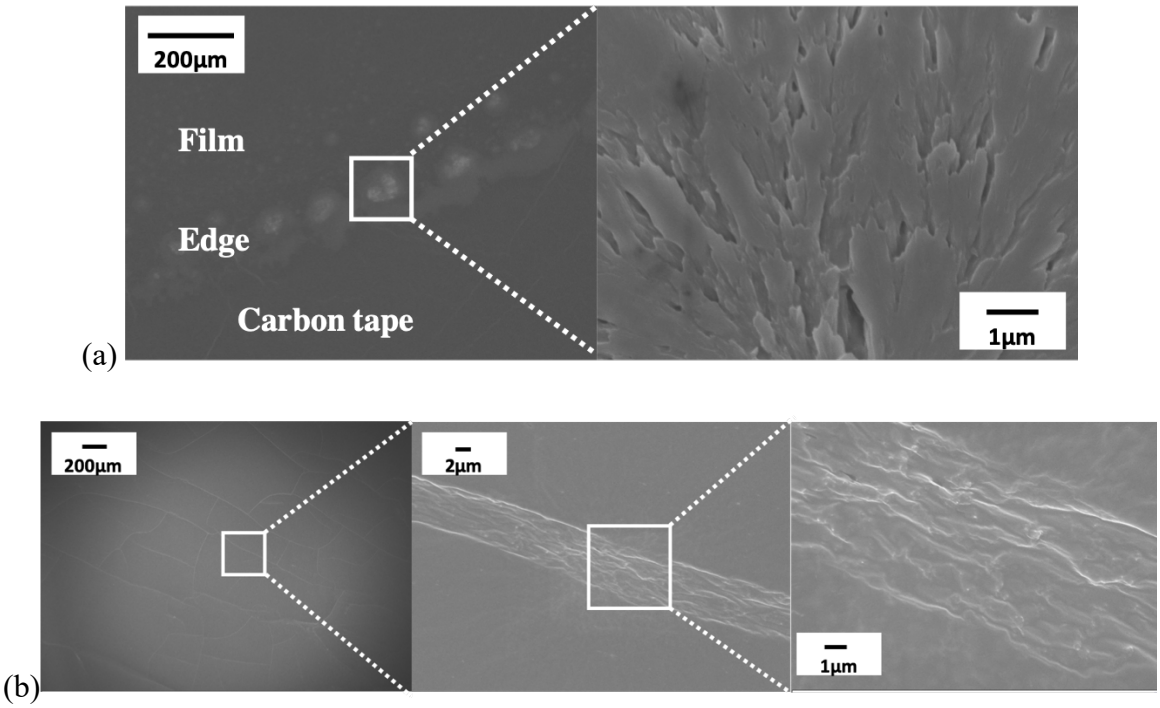


Figure 5.2– SEM images of large scale and zoomed in mcRFP protein film at 30 mN/m after compression.

Similarly, the mcRFP-Q77 film possesses a smooth surface at 7 mN/m (Figure 5.3 (a)) under SEM. The zoomed-in images both on the surface of the film and at the bottom of a crack in the film also show layered structure, similar to mcRFP. Compression of the mcRFP-Q77 film to 30 mN/m resulted in the wrinkled structure. These wrinkles (Figure 5.3 (b)), compared with the ones in mcRFP film (Figure 5.2 (b)) were smoother and the texture from the top view was more aligned instead of unorganized. Both mcRFP-Q77 films remained fluorescent under fluorescent microscope. (Figure 5.4). It is shown that the assembly of both mcRFP and mcRFP-Q77 resulted in layered films without any aggregated "island" structures. The fact that such layered film remained fluorescent for mcRFP-Q77 not mcRFP may be explained as the more stretched mcRFP molecules assembled on the interface, which temporarily lost its fluorescent. However, this explanation needs to be further confirmed and cannot be concluded here. The compression of the films, on the other hand, resulted in wrinkled film with crowded molecules. If the explanation is true, the fluorescent is recovered during the compression process through the rearrangement of the molecules in the film.

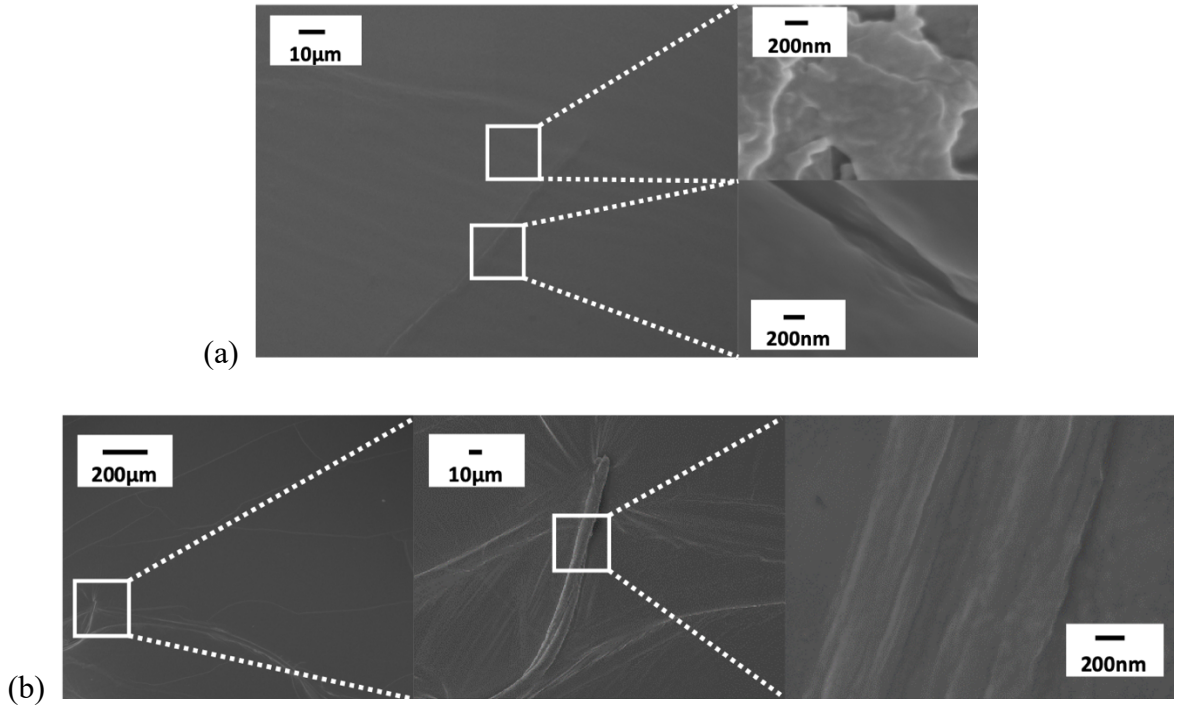


Figure 5.3– SEM images of large scale and zoomed in mcRFP-Q77 protein film (a) at 7 mN/m before compression and (b) 30 mN/m after compression. The two zoomed-in images in (a) were focused to different planes: the top one was focused to the surface of the film and the bottom one was focused to the bottom of the crack.

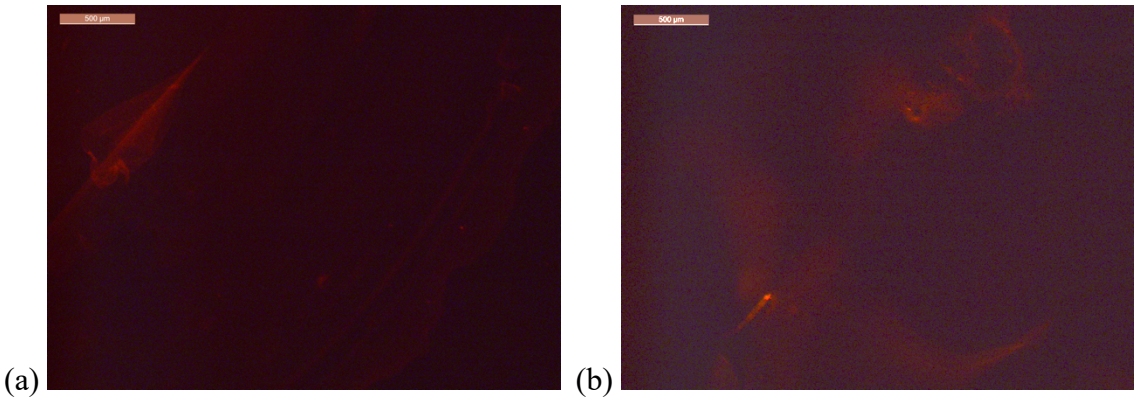


Figure 5.4– Fluorescent images of mcRFP-Q77 protein film at (a) 7 mN/m before compression and (b) 30 mN/m after compression. (Brightness of the images were adjusted for better visualisation, scale bars: 500 µm)

As is discussed in Chapter 3, a sharp peak in the FT-IR spectrum at 1652 cm^{-1} indicating the α-helical structure was observed with mcRFP-Q77 protein, whereas only a small shoulder at 1658 cm^{-1} was detected with mcRFP protein revealing a smaller proportion of the same secondary structural composition. In contrast, the IR spectra of both proteins in the film state resulted in a broad, featureless peak due to the strong overlap in amide I region (1600-1700 cm^{-1}). Similar to the self-assembled product from the last chapter, such a broad band might be indicative of the strong vibrational interactions among protein molecules in the film. The film is therefore considered to be a highly packed state of protein molecules. Figure 5.5 compared such broad amide I and amide II band of mcRFP and mcRFP-Q77 film at 30 mN/m after compression. Due to the well-controlled sample background, sample area, as well as the solid state of the film, the difference in the percentage of transmission could imply more proportion of protein molecules vibrating in the mcRFP-Q77 film, which might result from either a thicker or a more tightly packed film.

In order to get more information from the IR peaks, a deconvolution fit was performed for the spectra of both protein films.[232] The resulting peaks within the amide I region indicate a secondary structural change to β-sheet for both protein films. The mcRFP protein, which possess mostly β-sheet structure (intensive peak at 1628 cm^{-1}), retained most of the structure (intensive peak at 1633 cm^{-1}) with additional β-turns (1661 cm^{-1}) and some antiparallel (1679 cm^{-1}) structures reviewed in Figure 5.6 (a) after film formation. The results indicated a structural change during this process. With mcRFP-Q77 protein, equally intensive peaks were observed for β-sheet and α-helix structure (at 1632 and 1652 cm^{-1} respectively) before film formation; whereas the corresponding protein film showed only the β-sheet structure (1635 cm^{-1}) and an additional β-turns (1665 cm^{-1}) structure in the amide I region in Figure 3.9 (b). The sharp peak of α-helix appearing in mcRFP-Q77 protein was attributed to the structure of the polyQ tag. Therefore, the disappearance of this peak in mcRFP-Q77 film is an indication of the α-helix to β-sheet transition of the polyQ tag. This matched the hypothesis that the polyQ tag is prone to aggregate due to its β-sheet formation. And the development of H-bond among the side chain of glutamine repeats predominates over H-bond with water, thereby forming such an insoluble β-sheet rich product.

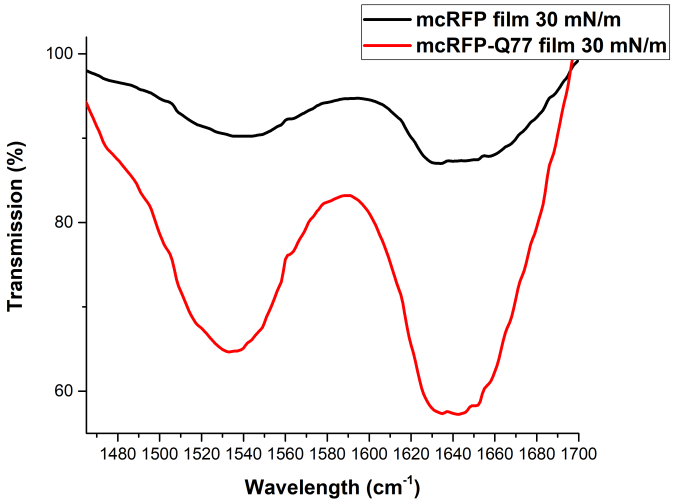
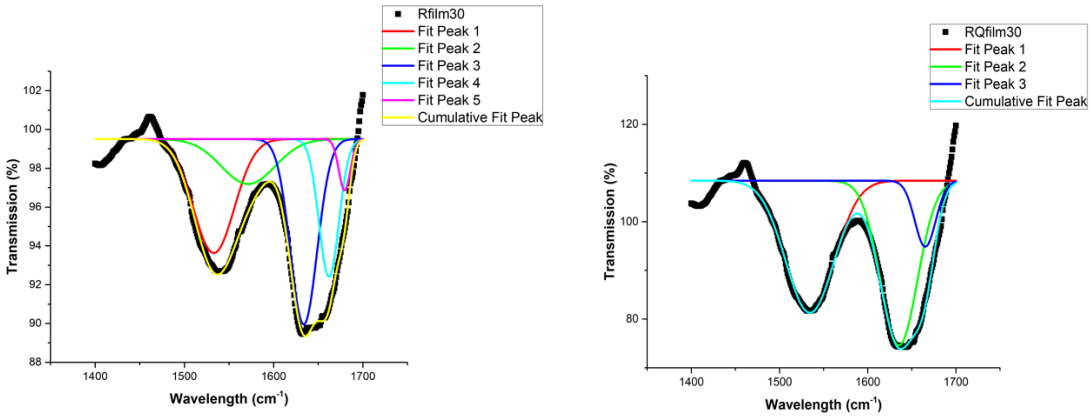


Figure 5.5– FT-IR spectra of mcRFP and mcRFP-Q77 film at 30 mN/m after compression.



(a) (b)

Figure 5.6– Deconvolution fit of FT-IR spectra of (a) mcRFP and (b) mcRFP-Q77 film at 30 mN/m after compression.

5.2.2 Protein Fibrous Materials

5.2.2.1 Three-Dimensional Structure of Protein Fibres

A 1-dimensional, self-supported protein fibrous material could be generated by lifting the protein film from the AWI. The lengths of these fibrous materials vary from 1 cm up to 20 cm (Figure 5.7 (a) and (b)); and the diameters of these fibrous materials range from 30 µm to 100 µm. The fibre length largely depends on the degree of protein assembly at the AWI, and hence the protein type. The longest fibres (Figure 5.7 (b)) were produced from the assembly of Luc-mcRFP. Preliminary characterisations under the fluorescent microscope give initial indication of the retention of the fluorescence activity upon fibre formation (Figure 5.7 (c) and (d)).

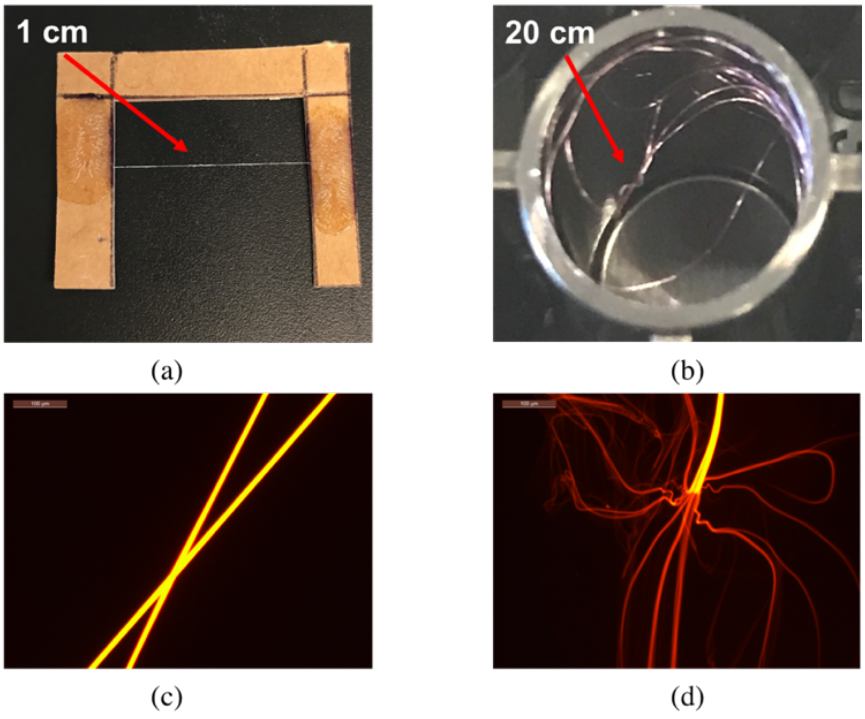


Figure 5.7– Images of mcRFP-Q77 (a) and Luc-mcRFP (b) fibres lifted from the AWI with different holders (length labelled), and the fluorescent microscope images of mcRFP fibres lifted from the AWI (c), (d). (scale bar: 100 µm)

SEM images demonstrate a morphological difference of mcRFP and mcRFP-Q77 tagged fibre from both the surface and cross-sectional view. The surface of the mcRFP fibre is randomly arranged with some hollow structure, and the wrinkles do not obtain any structured alignment (Figure 5.8 (a)). In comparison, the wrinkles on mcRFP-Q77 fibre are vertically aligned even at nanometre scale (Figure 5.9 (a)). These surface structures correspond well to the film structure of both proteins (Figure 5.2 (b) and Figure 5.3 (b)). Similarly, a randomly aligned sheet structure was observed with the cross-sectional view of mcRFP fibre (Figure 5.8 (b)); whereas a layer-by-layer sheet structure was found along the cross section of mcRFP-Q77 fibre. A zoom-in image for one of the sheets was shown in Figure 5.9 (b). Such a sheet cross sectional structure might be indicative of a layer-by-layer adsorption of mcRFP-Q77 at AWI. The zoom-in image of the sheet for both mcRFP and mcRFP-Q77 fibre shows a closely packed structure.

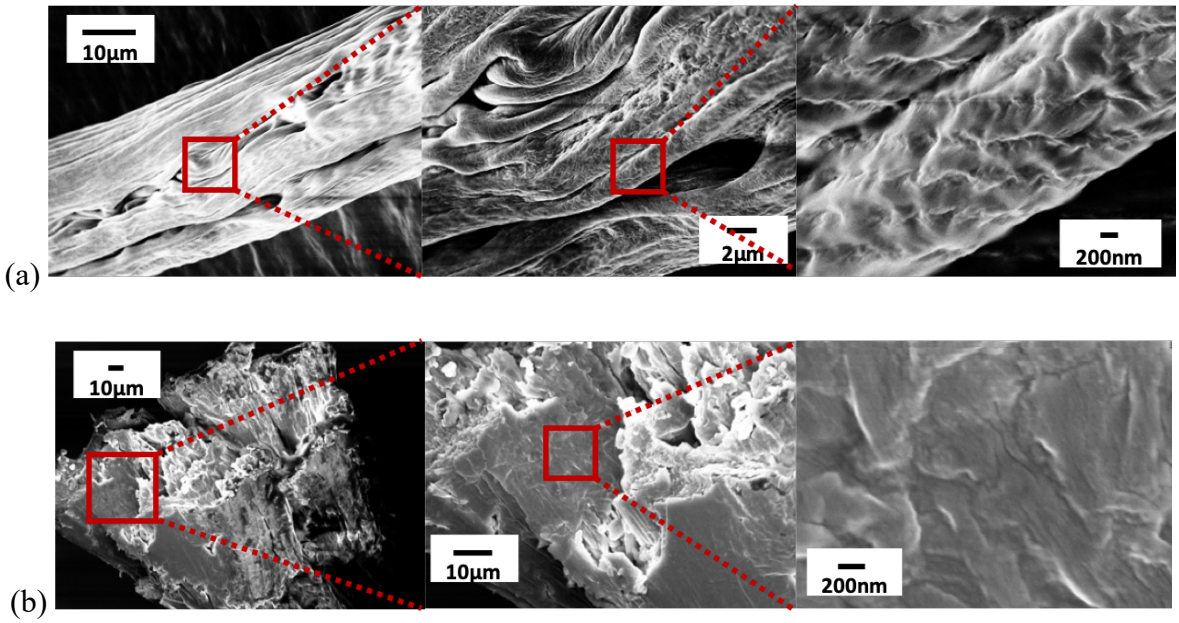


Figure 5.8– SEM images of mcRFP fibres from (a) the side view, and (b) a cross sectional view at different zoom-in scales.

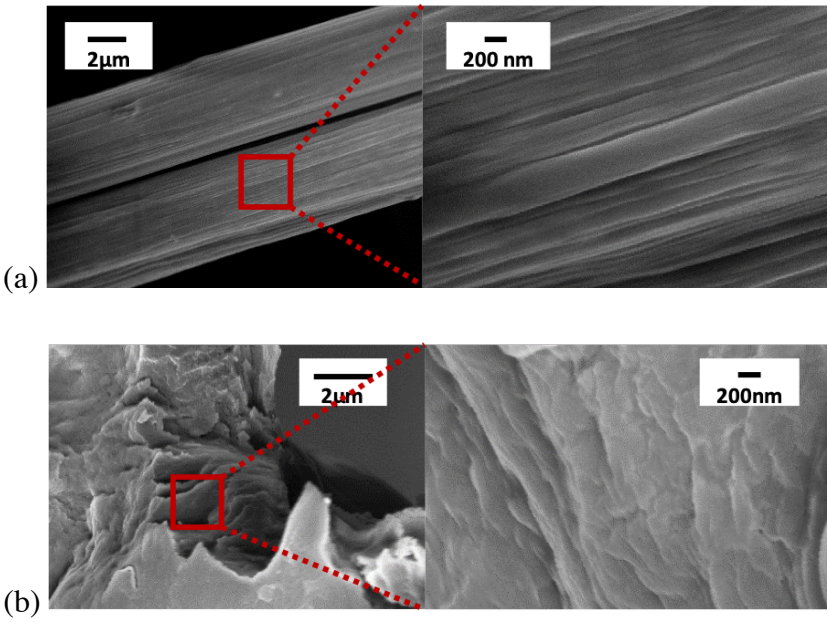


Figure 5.9– SEM images of mcRFP-Q77 fibres from (a) the side view, and (b) a cross sectional view at different zoom-in scales.

5.2.2.2 Mechanical Properties of Functional Protein Fibres

The layered cross-sectional structures of the mcRFP/mcRFP-Q77 fibres show much similarity with silk material as discussed in Chapter 1 (Figure 1.8). Such structure has been reported to affect the unique mechanical properties of silk.[233] Therefore, mechanical properties of the mcRFP/mcRFP-Q77 are examined. Table 5.1 compares the mechanical properties of the experimental carbon fibre, mcRFP/mcRFP-Q77 fibres, and the literature reported collagen fibres, native spider silk and bombyx mori silk fibres. It is shown here that mcRFP and mcRFP-Q77 fibres behave differently as they break. The modulus, breaking stress and strain of both protein fibres are in the same range with collagen and are much smaller than silk. Comparing only the two protein fibres produced, Q77 largely increased the mechanical strength. It is noticed that there is a 3-fold increase in the mechanical strength when mcRFP is tagged with Q77. This is indicative of increased β-sheet structure formed with intramolecular H-bonding and intermolecular Q77 stacking. These increased hydrophobic interactions strengthened protein fibres with Q77 tagged structures. Unlike the triple helical structure of collagen which

prevents the elongation of the fibre, mcRFP and mcRFP-Q77 fibre exhibits silk-like cross sectional structures. Even though the breaking strain of both fibres are much smaller than silk, mcRFP exhibits a superior extensibility than collagen. It was previously reported that the elasticity and elongation strain of silk is affected by the H-bond in the system.[234] Therefore, the larger elongation strain of mcRFP could also support the H-bond formed within the polyQ structure, which makes mcRFP-Q77 even harder to stretch under external forces. Traditionally, the extensibility of a disease-related peptide fibre may not be experimentally tested due to their nanoscale sizes. This is the first time that the extensibility of a disease-related protein fibre is experimentally examined due to the special production method from an AWI for this Q77 tagged protein fibre into centimetres scale.

Table 5.1 – Mechanical properties of carbon fibre, fibres of mcRFP constructs, and native silk fibres.

	Stress (MPa)	Strain (%)	Modulus (GPa)
Carbon fibre	689 ± 17	0.22 ± 0.02	80.7 ± 1.3
mcRFP fibre	16 ± 4	4.5 ± 2.2	0.3 ± 0.2
mcRFP-Q77 fibre	20 ± 8	0.6 ± 0.3	1.1 ± 0.6
Collagen fibre from animal bone[235]	63 ± 14	2.2 ± 0.41	1 ± 0.4
Spider dragline silk from *Nephila Claipes*[236]	1030 ± 110	18 ± 6	13.7 ± 3
Silkworm silk from *Bombyx Mori*[237]	398 ± 51	27 ± 0.8	9.8 ± 0.6

Necking is a mode of tensile deformation where relatively large amounts of strain localize disproportionately in a small region of the material.[238] This behaviour is often associated with

ductile materials such as metals and polymers, resulting in a yielding in the stress-strain curve. Figure 5.10 explains the different behaviour of the protein fibre compared with carbon fibre which breaks as soon as the force reaches the breaking point of the material. It is observed that mcRFP fibre exhibited a long stable necking period, which could also be captured under the microscope with prominent decrease in the size of local cross-sectional area. On one hand, the longer necking time of mcRFP fibre compared with the mcRFP-Q77 fibre can be explained by the difference of their cross-sectional structure. The interactions between horizontally aligned layers of mcRFP-Q77 (Figure 5.9 (b)) might be stronger than the randomly aligned mcRFP layers. And the force is more evenly distributed on mcRFP-Q77 fibre. On the other hand, this behaviour can be explained as a modular elongation mechanism by Smith et al.[239] This mechanism described that natural fibres experience modular unfolding until no folded domains are left before the fibre breaks. This might be used to explain the green to yellow colour change found in the mcRFP fibre necking region, and the yellow colour found in the mcRFP fibre breaking region. This change in colour might be indicative of a change in chromophore folding structure, resulting in a reduction of the red chromophore back to the green intermediate with shorter π-conjugation (Figure 1.4).

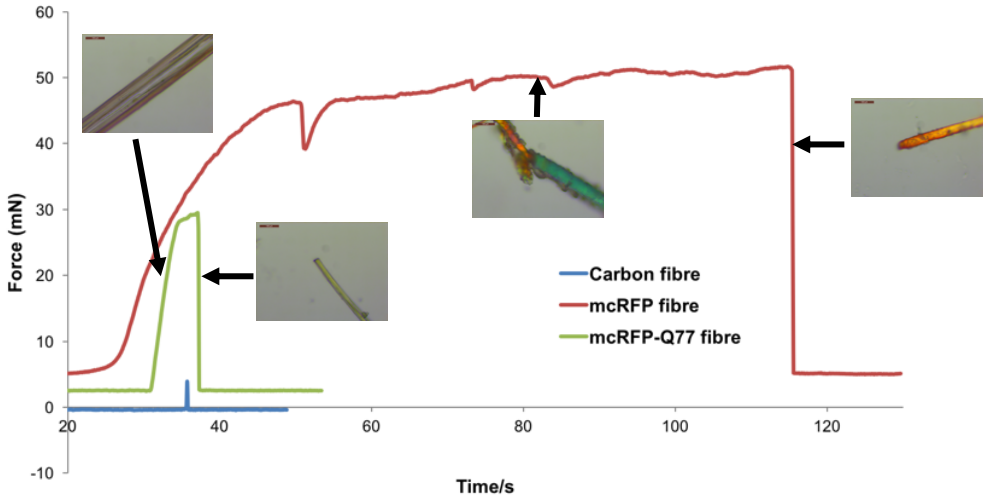


Figure 5.10– Stretching behaviour of fibres with inserted microscopic images corresponding to the behaviour (Scale bars: 100 μm).

5.2.3 Functional Integrity of Protein Fibres

5.2.3.1 Fluorescence of Recombinant mcRFP Fibres

In order to quantify the fluorescence in protein fibres, calibration curves of both native mcRFP and mcRFP-Q77 solution were made by recording the fluorescence at each concentration under a home-built BiognostiX reader. This method has a sensitivity down to $3x10^4$ calculated pixel intensities, which corresponds to 20 pmoles of native mcRFP and 50 pmoles of native mcRFP-Q77 in solution. The detection limit is much lower than a fluorometer, which is in nmoles range of mcRFP. The total intensity obtained for mcRFP-Q77 is 2.2 times smaller than that for mcRFP, calculated from the slope of the linear range intensity Appendix A.6 Figure A6.7. This is supported by the result measured with the commercial fluorimeter, that the fluorescent intensity is 2.4 times smaller (40% activity retention) for mcRFP-Q77 than mcRFP (Figure 3.3 (a)), validating the success of using this method to calculate the total intensity. Similarly, the calibration curves of both Luc-mcRFP and Luc-mcRFP-Q77 were obtained (Appendix A.6 Figure A6.8). The linear fits with intercepts indicate the detection limits equivalent to 45 and 55 pmoles of Luc-mcRFP and Luc-mcRFP-Q77, respectively. These calibration curves are used to quantify the fluorescence retained in terms of the equivalent number of native state molecules in each type of fibre.

Protein assays are widely used to determine the total amount of protein in a certain sample solution, which may be utilised when both the standards and the fibre samples are dissolved in the solution. It has been shown that both the mcRFP and mcRFP-Q77 fibres are soluble with treatment and incubation in 8 M urea. Therefore, the calibration curve is obtained with urea denatured mcRFP and mcRFP-Q77 solution (Appendix A.6 Figure A6.9).

The fluorescence of protein fibres was first recorded under the BiognostiX reader. By comparing with the fluorescence of corresponding solution, the equivalent number of molecules which remains fluorescent could be obtained. The fibres were then dissolved in 8 M urea before they were mixed with the BCA reagents. The absorbance of the mixture at 562 nm was compared with the calibration curve of the corresponding urea denatured protein solution to get the total number of protein molecules in the fibre.

A plot of fluorescent molecules versus total protein molecules (Figure 5.11) was obtained for both mcRFP and mcRFP-Q77 fibres. A general trend of increasing equivalent fluorescent molecules with increased amount of total protein molecules in each type of the fibre was observed. The initial flat region is indicating the equivalent number of fluorescent molecules in the fibre is below the detection limit for each type of protein using this BiognostiX reader method. It shows that mcRFP in the fibre retained 18.3% ± 4.9% of its total fluorescence in its native state, whereas 74% ± 13.4% fluorescence was retained for mcRFP-Q77 when it forms fibres. As is shown in Figure 5.11, with the same amount of protein in a fibre, there are more equivalent fluorescent molecules in the mcRFP-Q77 fibres, compared to the mcRFP fibre. Considering that 40% of mcRFP activity was retained upon Q77 fusion, the final retention of fluorescence in mcRFP-Q77 fibre is 29.6%, compared to 18.3% in mcRFP fibre (Table 5.2). The results indicate that during the protein assembly at AWI, the polyQ tag, to a certain extent, prevents the unfolding of mcRFP by rearranging the secondary structure of its own helical conformation. This explanation matches the conclusion from the last chapter, that the polyQ tag adopted a β-sheet structure via intramolecular H-bonding at the AWI, as also suggested by the IR spectrum. The mcRFP folding structure, and hence its chromophore, was consequently protected from deformation and retained a higher fluorescence with the corresponding fibre product with this polyQ tag.

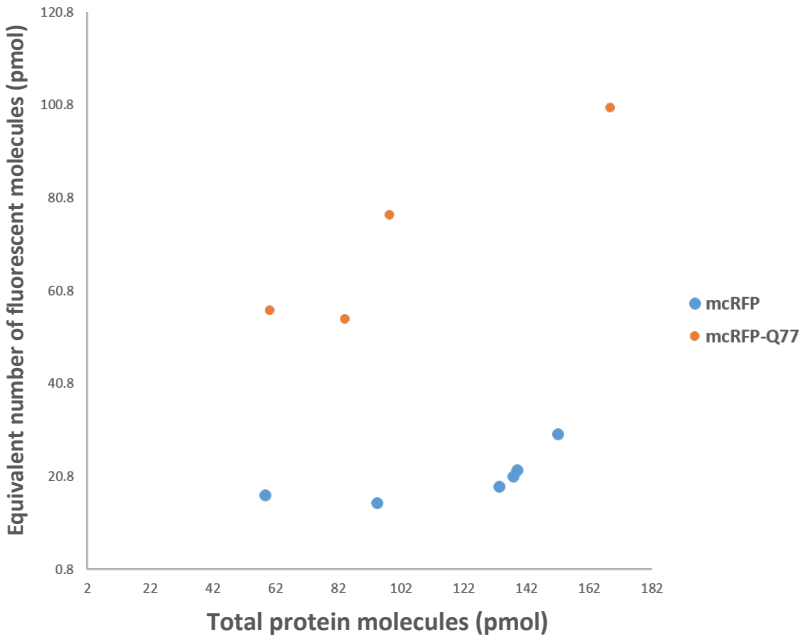


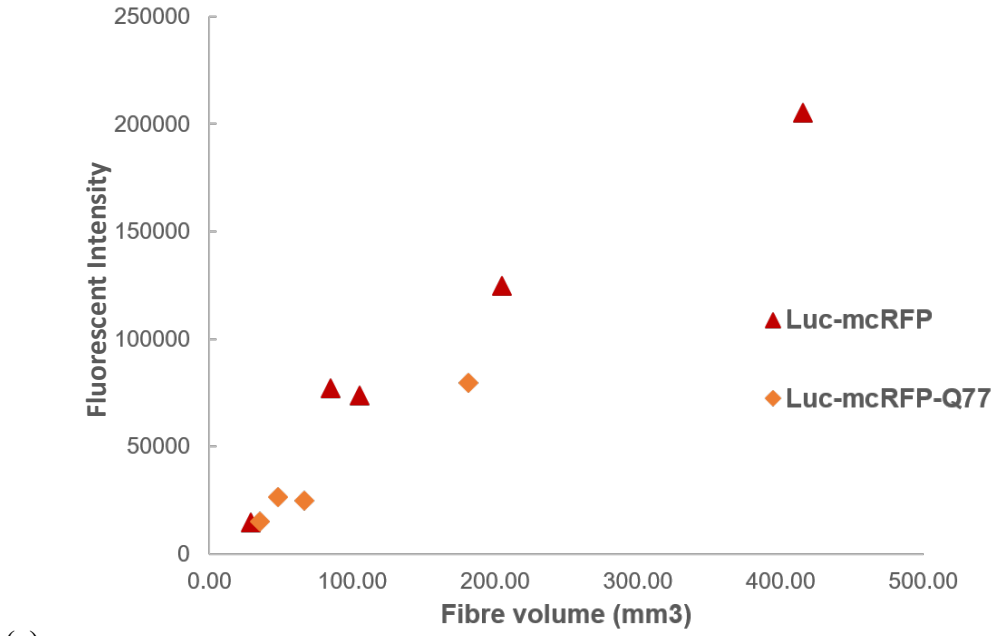
Figure 5.11– Fluorescence retention in (a) mcRFP fibre and (b) mcRFP-Q77 fibre.

Table 5.2 – Comparing fluorescence retention in recombinant mcRFP-Q77 solution, mcRFP fibre and mcRFP-Q77 fibre.

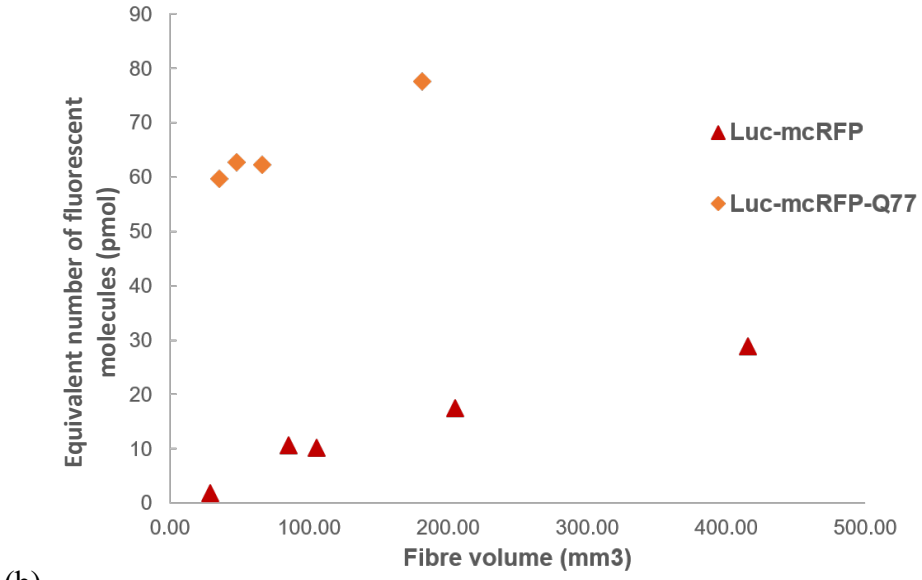
Protein	Fluorescence retention from native mcRFP (%)
mcRFP-Q77	40.1 ± 2.5
mcRFP fibre	18.3 ± 4.9
mcRFP-Q77 fibre	29.6 ± 5.4

With the pair of Luc-mcRFP, the protein fibres were quantified with the calculated fibre volume due to a visually lower solubility of Luc-mcRFP fibres. The fibre fragments were still noticeable in 8 M urea after 2 hours of incubation and pipette mixing. In this case, the percentage of fluorescence retained cannot be calculated due to the unknown total number of molecules in the fibre. However, it is assumed that the packing density in the fibre drawn from the same film remains the same. Therefore, the total number of molecules should be proportional to the volume of the fibre. For both Luc-mcRFP and Luc-mcRFP-Q77, it is

noticed that the equivalent number of fluorescent molecules increased with increasing fibre volume. It is found that the fluorescent intensity emitted from the same volume of Luc-mcRFP fibre and Luc-mcRFP-Q77 fibre were nearly the same (Figure 5.12 (a)). Taking the original fluorescence into consideration, that the fluorescent intensity of native Luc-mcRFP is 4 times higher than native Luc-mcRFP-Q77 (Figure 3.3 (b)), it is concluded that Luc-mcRFP-Q77 retained more fluorescence than Luc-mcRFP when they assemble at the AWI and forms the fibres. This conclusion is seen in Figure 5.12 (b), that Luc-mcRFP-Q77 retains a larger number of fluorescent molecules, comparing to the same volume of Luc-mcRFP fibre. This may be explained that even though a slower assembly was found for Luc-mcRFP-Q77 comparing to Luc-mcRFP, further conformational changes at the interface before and after compression may expose the spatially protected Q77 to interact with each other, and thus protect the mcRFP domains from interactions. Therefore, smaller conformational change happened in the mcRFP domain of Luc-mcRFP-Q77.

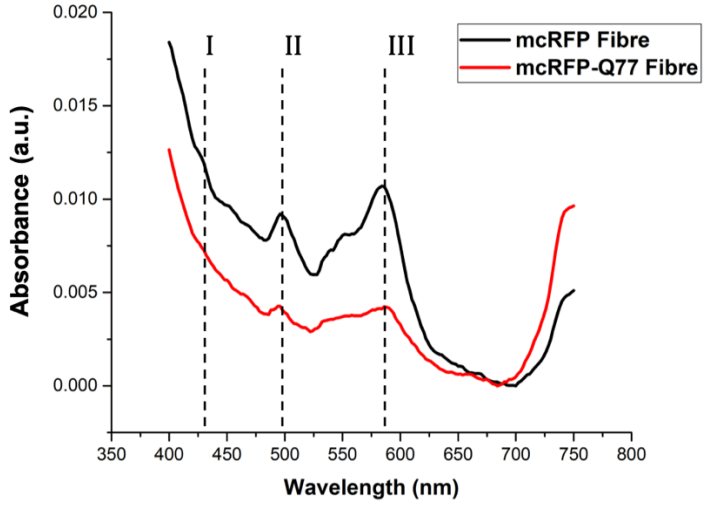


(a)

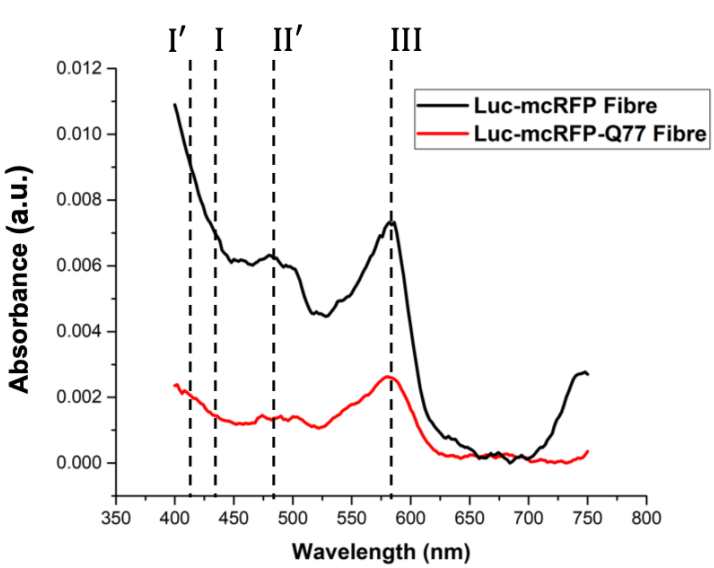


(b)

Figure 5.12– Fluorescence retention of Luc-mcRFP fibre and Luc-mcRFP-Q77 fibre. (a) Fluorescent intensity under BiognostiX reader; (b) equivalent number of corresponding fluorescent molecules in their native states retained in each fibre and (b) absolute total intensity under BiognostiX reader.



(a)



(b)

Figure 5.13– Absorption spectra of (a) mcRFP and mcRFP-Q77 fibres, and (b) Luc-mcRFP and Luc-mcRFP-Q77 fibres. The absorption peaks are I' and I: blue at 408 nm and 426 nm, II' and II: green at 486 and 497 nm, and III: red at 586 nm.

Apart from the relative quantification of the fluorescent activity obtained with the fibres, determination of chromophore properties could provide further information on protein structural change during fibre formation. Figure 5.13 demonstrates the absorption spectra of the mcRFP contained fibres. No shifts in the peak position of the red chromophore was

observed (586 nm) for all the four constructs. For the mcRFP pair (Figure 5.13 (a)), the absorption of the green chromophore experienced a slight blue shift of 5 nm (from 502 nm shifted to 497 nm), and the blue chromophore experienced a large red shift of 16 nm for both mcRFP and mcRPF-Q77 (from 408 nm shifted to 426 nm). This is resulted from the post-translational modification on the chromophore, which experienced a change in the microenvironment. The large red shift might be indicative of the stack interaction of the blue chromophore; whereas the blue shift is mainly caused by a positively charged chromophore, or a less polarised double bond for the green intermediate in the fibre. However, for Luc-mcRFP (Figure 5.13 (b)), the peak position of the blue chromophore also experienced a red shift from 408 nm to 426 nm, but that for Luc-mcRFP-Q77 remained at 408 nm. The green chromophore in both protein fibres broadened and with the peak centre at 486 nm. This larger blue shift might be indicative of the change in the surrounding amino acid of the green chromophore, which experienced a less efficient deprotonation process, and a less polarised double bond in the fibre.[92]

Figure 5.14 illustrates the proportional change of the three coloured chromophores in each type of protein before and after the formation of fibre. As is explained in Figure 3.4 (b)- the absorbance proportion of different chromophores, all the three chromophores exist during the mcRFP maturation have similar extinction coefficient. Therefore, the proportion of each chromophore could be compared from the total absorption. The proportion of the red chromophore decreases after fibre formation for all constructs (Figure 5.14 (a)). The green chromophores also occupy a smaller percentage in all types of proteins after fibre formation (Figure 5.14 (b)), and the reverse for the blue chromophores (Figure 5.14 (c)). These changes in the green and blue chromophores both indicate the higher energy barrier for the deprotonation process to occur during unfolding and refolding steps of the fibre formation, which may be due to a "twist" in the bond with the cis-trans isomeric transformation in the chromophore (Figure 1.4).

Despite the general trend for all constructs, it is noticed that the Luc-mcRFP pair experienced the larger decrease in the proportion of the green chromophore, compared to the mcRFP pair, indicating largely decreased in the deprotonation process towards the green intermediate. It is interesting to find that mcRFP-Q77 and Luc-mcRFP have the largest increase in the blue

chromophore proportion, which is indicative of the largely decreased second deprotonation process for these two proteins. This change in the second deprotonation process also matches the protein assembly behaviour at the AWI. Both mcRFP-Q77 and Luc-mcRFP assemble faster than their partners, mcRFP and Luc-mcRFP-Q77, respectively. The results might be indicative of a stronger H-bonded intramolecular interaction for mcRFP-Q77 and Luc-mcRFP, which both prevents the deprotonation process at a larger degree and facilitates the initial assembly at the AWI.

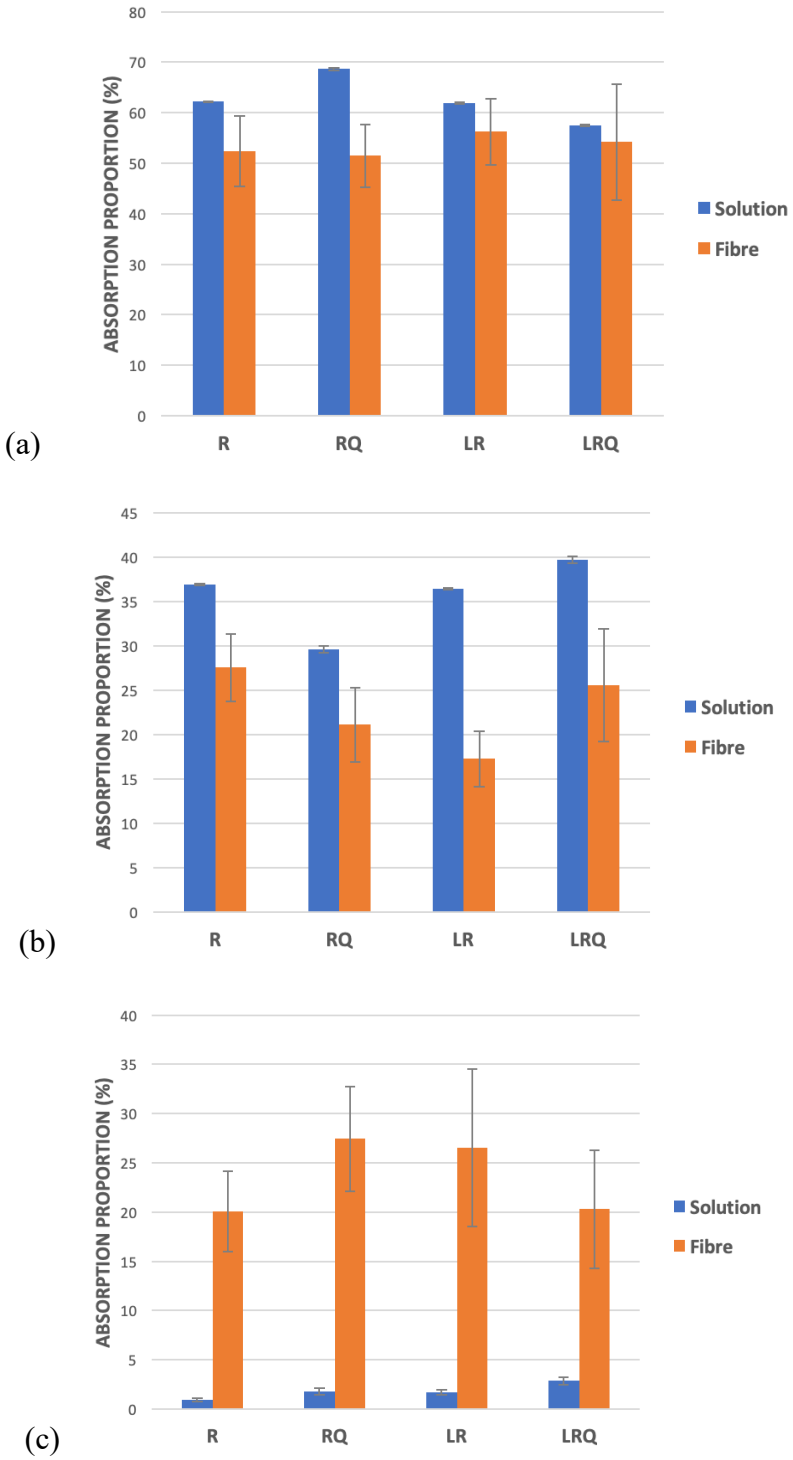


Figure 5.14– Relative percentage absorption of (a) red (b) green (c) blue chromophore for mcRFP contained protein in solution and in fibre.

Table 5.3 further compare the fluorescent lifetime of the mcRFP pair and the Luc-mcRFP pair after fibre formation, which indicates solvent accessibility changes of the chromophore after fibre formation. Comparing with Table 3.3, the results show a general increase of the lifetime in the blue chromophore after fibre formation apart from mcRFP. In contrast, the lifetime for the green and red chromophore decreases for all proteins apart from mcRFP-Q77. This is indicative of an increased solvent accessibility of the chromophores in mcRFP. With other constructs, solvent accessibility towards the blue chromophore is decreased, which might be caused by the tightly packed blue chromophore through electrostatic charge as it carries both a negative and a positive charge.

Table 5.3 – Fluorescent lifetime of mcRFP contained protein molecules in fibres.

Protein Fibre		mcRFP	mcRFP-Q77	Luc-mcRFP	Luc-mcRFP-Q77
Lifetime (ns)	Blue	0.14	0.19	0.18	0.15
	Green	1.62	4.34	1	1.76
	Red	2.93	4.34	1.83	1.76

5.2.3.2 Bioluminescence of Recombinant Luciferase Fibres

With the β-barrel protected chromophore, mcRFP retains nearly 20% activity after fibre formation. The addition of the polyQ tag helps the recovery of the fluorescence of the fibre due to its self-assembly property at the air-water interface. As was discussed in the last chapter, Luc behaves differently at AWI compared to mcRFP. Therefore, it is interesting to compare the fibre formation of this enzymatically active protein with the reporter mcRFP, in order to achieve the aim of a real functional protein fibre for sensing application.

The bioluminescent activity of luciferase fibres and luciferase-Q77 fibres was compared in terms of the total volume of the fibre. The results, as for the BCA assay detected protein amount in mcRFP and mcRFP-Q77 fibres, show a general increasing trend with increasing fibre volume (Figure 5.15). Unexpectedly, the production of a long enough fibre for Luc-Q77 failed. Only small fibres with volumes up to 50 mm^3 were produced. This might indicate a different status of the protein film for Luc-Q77. As was discussed in the previous chapter, compression of Luc-Q77 at 30 nmoles, 10 μM, resulted in a lower surface pressure, which might be indicative of the less β-sheet structure formed (Figure 4.12 (b) black line). Therefore, the film was less self-supportive, and the length of the fibre drawn from the film was limited. At a higher concentration, even though the polyQ tag facilitated the assembly at AWI, the final surface pressure after compression was low. The compressibility (slope in the **Π**-A isotherm) was small. This slower increase in the surface pressure might indicate a limited amount of intramolecular linked protein molecules in the film. And the desorption effect is significant during the compression. Therefore, the length of the fibre produced was limited.

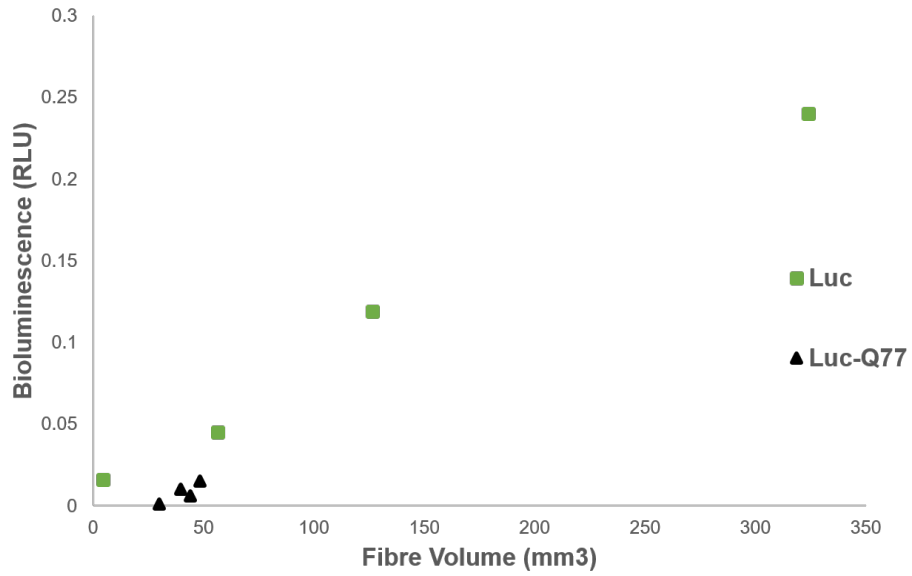


Figure 5.15– Bioluminescent activity in luciferase and luciferase-Q77 fibres.

The Luc-mcRFP and Luc-mcRFP-Q77 proteins benefit from their dual funcationalities. Figure 5.16 compares the equivalent number of luciferase active molecules in these two protein fibres, according to the calibration curves shown in Appendix A.6 Figure A6.10. Like the mcRFP

active molecules, the number of Luc active molecules were increasing with increased fibre volumes. The equivalent number of corresponding native state luciferase activity remained the same both with and without the Q77 tag, showing no effect of the Q77 for luciferase activity in this pair of fibre product. In theory, the number of luciferase molecules and mcRFP molecules in each fibre should be the same for both Luc-mcRFP and Luc-mcRFP-Q77 constructs. However, for both constructs, the number of Luc active molecules were 3 orders of magnitude less than the mcRFP active molecules (fmoles range versus pmoles range comparing Figure 5.16 and Figure 5.12). It can then be concluded that, during fibre formation at the AWI, the mcRFP retained more activity compared to the luciferase which is highly affected by even minor conformational change due to the active site located in between the two mobile domains. It was discussed in previous chapters that the addition of Q77 spatially separated the interaction between Luc and mcRFP (Figure 3.13 (d) (e)), which slew down the rate of assembly (Figure 4.11 (b)). Therefore, the dominating part of assembly in both proteins, should be the molecular larger luciferase molecules. And the partially unfolded Luc in native Luc-mcRFP assembled slightly faster than Luc-mcRFP-Q77.

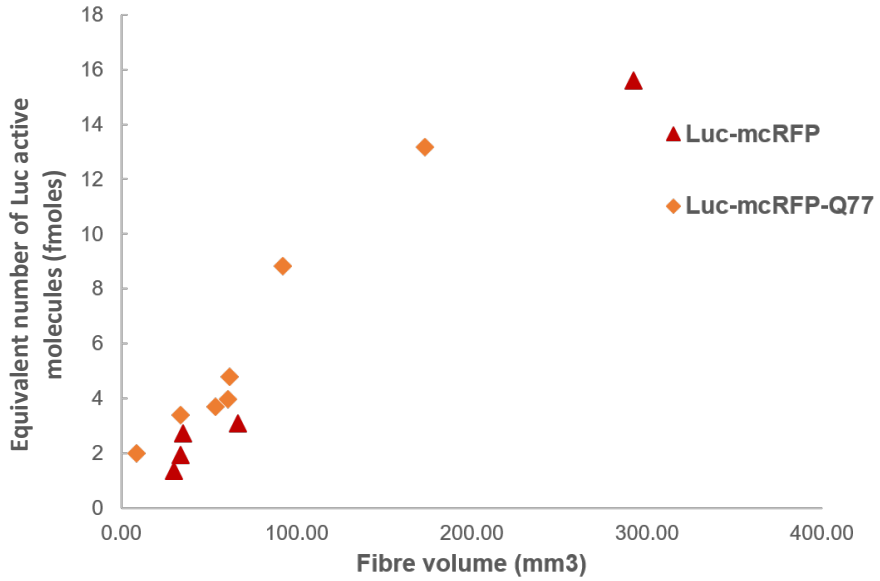


Figure 5.16– Luciferase activity in terms of equivalent number of native state luciferase active molecules in Luc-mcRFP and Luc-mcRFP-Q77 fibres.

5.2.3.3 Reaction Kinetics of Protein Fibres

The traditional BCA test was carried out with the calibration from the soluble BSA, in order to investigate the inhomogeneous reaction between the fibre and the solution. The results show clearly that the calibration with a soluble protein is not suitable for the determination of the total number of molecules in this solid fibrous structure (Figure 5.17). The reason for an inconsistent relationship between the measurement of total number of protein molecules in a fibre via BCA and the number of mcRFP/luciferase active protein molecules may be explained in Figure 5.18. When the BCA assay was performed with protein in a solution (Figure 5.18 (a)), each protein molecule is readily reacting with the environmental cupric molecules. As the reaction was incubated at 37 °C for 30 minutes, more and more cupric ion was reduced by the protein, forming cuprous cation chelated with BCA, which gives a purple colour. However, in the fibre system (Figure 5.18 (b)), extra time should be allowed for dissolving the fibre. In this case, diffusion time should be added into consideration. As the fibre could be considered as a "highly concentrated solution", which cannot be mixed with the BCA solution, a large amount of Cu^{2+} molecules near the fibre are consumed, resulting in a diffusion layer depletion and a

decreased reaction rate. Furthermore, the BCA assay was technically difficult for the fibre already being tested in luciferase assay. Unlike the transfer from a glass substrate to the assay solution, such solution to solution transfer causes fibre loss during transfer.

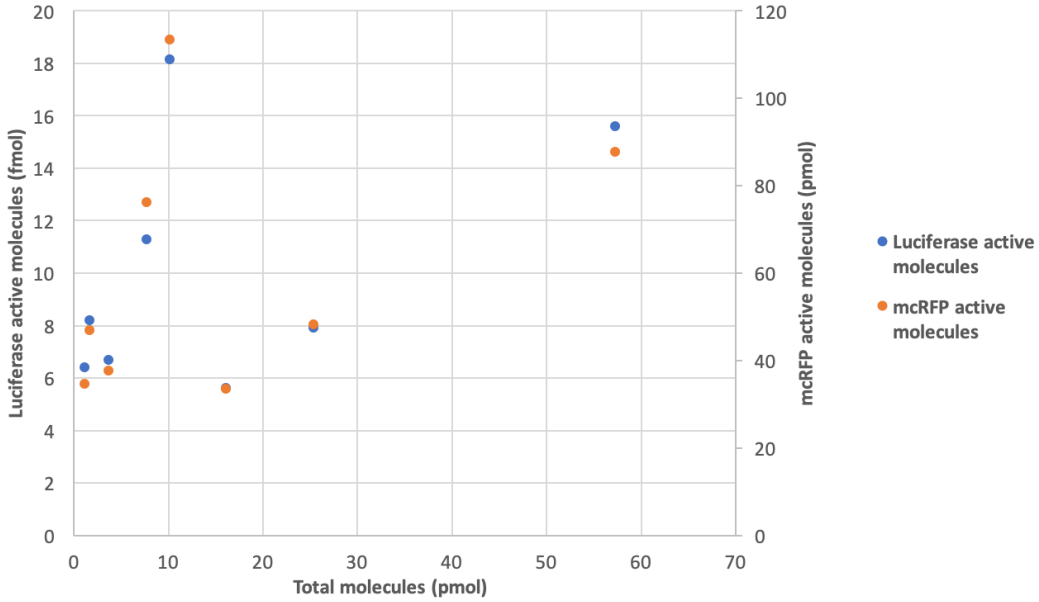


Figure 5.17– Plot of equivalent number of luciferase active molecules in fmoles range and mcRFP active molecules in pmoles range over total number of protein molecules in Luc-mcRFP fibre.

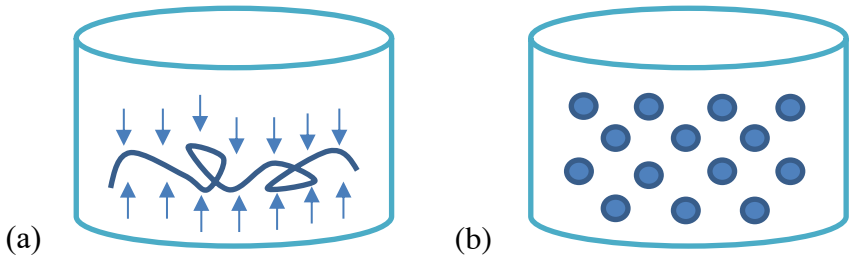


Figure 5.18– Illustration of reaction scheme of BCA protein assay in (a) the fibre system and (b) the solution system.

In contrast, a linear relationship between the equivalent number of mcRFP active molecules and Luc active molecules in the Luc-mcRFP fibre was noticed with 97% confidence (Figure

5.19). These numbers of active molecules were calculated from the corresponding bioluminescence and fluorescence calibration curves (Appendix A.6 Figure A6.10 and Figure A6.8 respectively), which were obtained using the Luminometer and the BiognostiX reader, respectively. The equivalent number of luciferase active molecules were found to be roughly 6000 times smaller than the number of mcRFP active molecules, which in theory should be equal to each other. The results showed a better retention of the stable mcRFP than the structurally flexible Luc. Interestingly, the light emission from Luc-mcRFP catalysed reaction was stable after the maximum was reached (Figure 5.20) without further decrease resulting from the competitive inhibition by the reaction product, oxyluciferin and L-AMP, as is found for the native Luc protein solution catalysed reaction. Even though it is hard to analyse this fibre catalysed reaction through the substrate affinity information, due to the uncontrollable production of the fibre by this current procudure, it might be deduced by the flat region of the maximum light emission, that the binding affinity of both the substrate and the inhibitor might be smaller. As expected in the inhomogeneous reaction, if the reaction is fast enough, a depletion layer around the fibre will be formed due to the consumption of the substrate and the production of the substrate. At the point of the maximum light emission, the diffusion reaches equilibrium. When the binding of the inhibitor is slower than the diffusion of the bulk substrate towards the fibre, the reaction could carry on until all the substrate was consumed in the solution. Therefore, the time for maximum light emission was largely extended compared to the native enzyme system. It is also shown in the bioluminescence profile that there was an increase in the lag time and a longer growing period towards maximum light emission. The maximum enzyme turnover, however, could still reflect the total number of active enzyme molecules in the fibre. The initial lag time of native state luciferase proteins catalysed reaction appeared to be uniform regardless of the protein type (Figure 3.6 (a)) and protein concentration (Figure 5.20). With luciferase-mcRFP solution, for instance, the lag time was tested to be 0.1 sec. This lag time was caused by the binding of luciferin with the enzyme. The reaction reached maximum light emission in about 0.5 sec, which corresponds to the conformational change of the protein. The formation of the fibre increases the time length in both periods. The delay in the time for the reaction to reach maximum light emission was not uniform even for the same type of luciferase fibres (Figure 5.20). This is explained by the variation of each fibre produced.

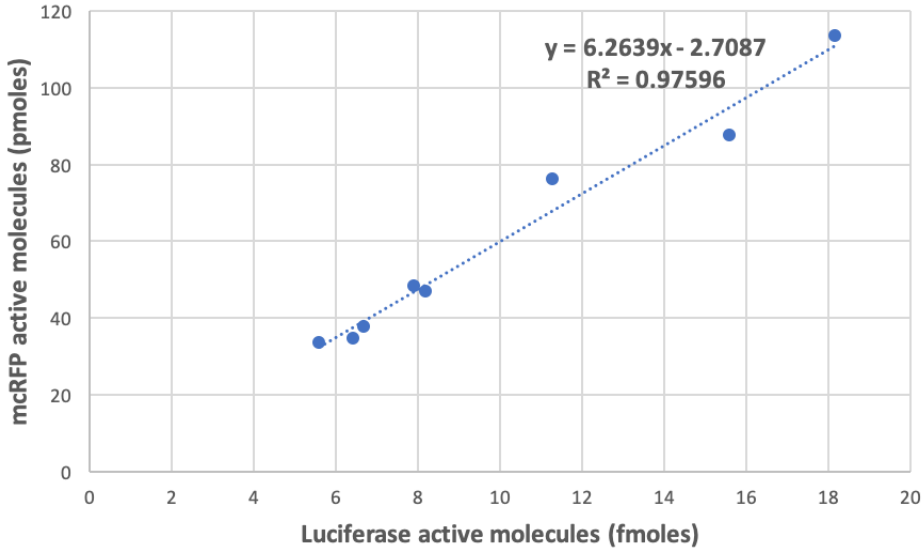


Figure 5.19– Plot of equivalent number of luciferase active molecules in fmoles range versus mcRFP active molecules in pmoles range in Luc-mcRFP fibre.

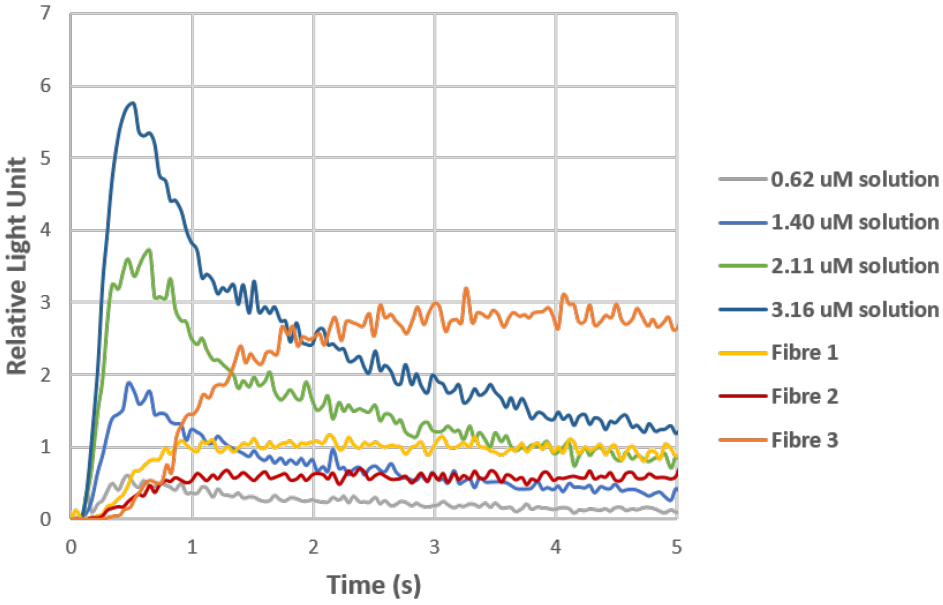
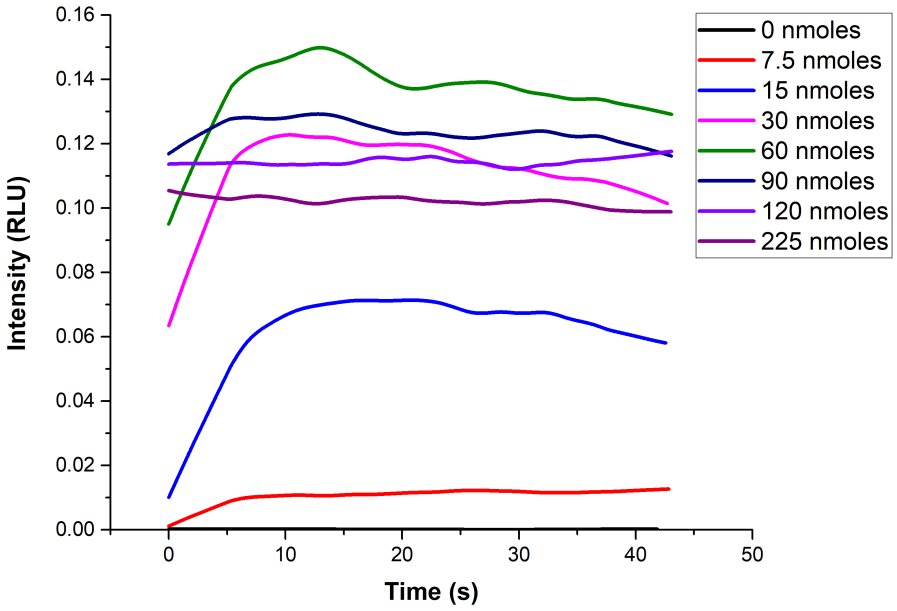
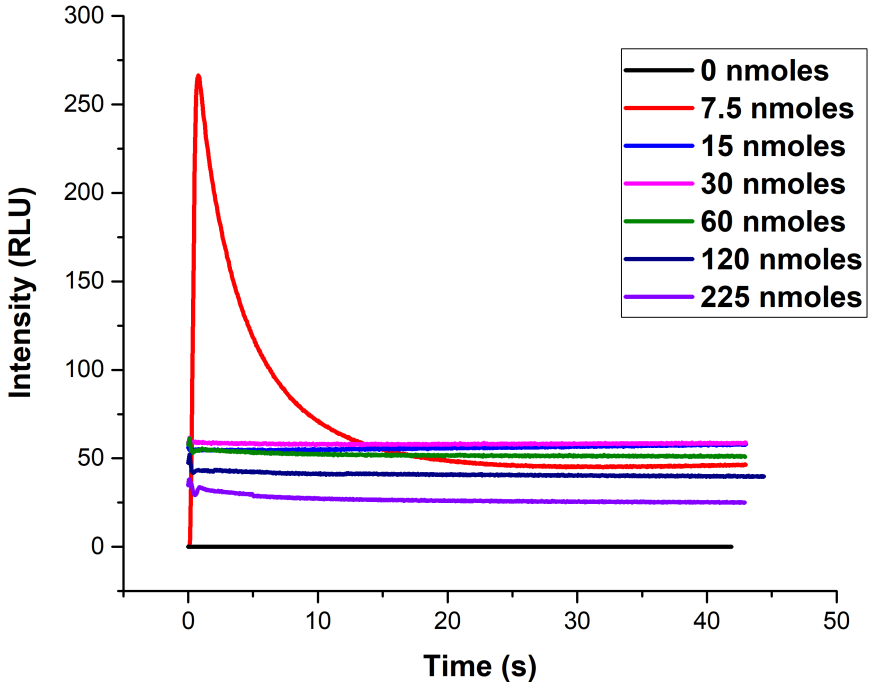


Figure 5.20– Kinetics of bioluminescence reaction with Luc-mcRFP solution at different concentration and Luc-mcRFP fibres produced from different batches.

Inspired by this non-decaying property of the light emission, an ATP assay was developed with the Luc-mcRFP fibre. ATP was added gradually to the system containing the enzyme and 200 μM D-LH_2 in 1x TEM buffer. Figure 5.21 compares the bioluminescent profile for a piece of Luc-mcRFP fibre and 1 pmole of Luc-mcRFP in solution. As was discussed above, the luciferase active molecules in the fibre are within fmoles range, so from the value of the maximum bioluminescent intensity (0.15 RLU, compared with the 270 RLU for the 1 pmole LR solution), it is confirmed that ATP was in excess even at the smallest concentration (7.5 nmoles) and so will not be completely consumed by the enzyme during the assay. However, the decay of the light intensity at concentrations higher than 15 nmoles was possibly indicating that the D-LH_2 had been consumed and was limiting the reaction, due to the long period of reaction. The plot indicates an increase in the light emission with increased ATP concentration catalysed by the Luc-mcRFP fibre, compared with the protein in solution. The integrated light emission at each ATP concentration was plotted in Figure 5.22. The preliminary result shows the ATP sensing capability of Luc-mcRFP fibre. Further variation of the ATP concentrations will determine the limit of detection for this fibrous system. The result demonstrated the first possibility of ATP sensing with the Luc-mcRFP dual functional protein fibre.



(a)



(b)

Figure 5.21– Bioluminescence intensity over time with ATP titration for (a) Luc-mcRFP fibre, and (b) Luc-mcRFP solution in 200 μM D-LH_2.

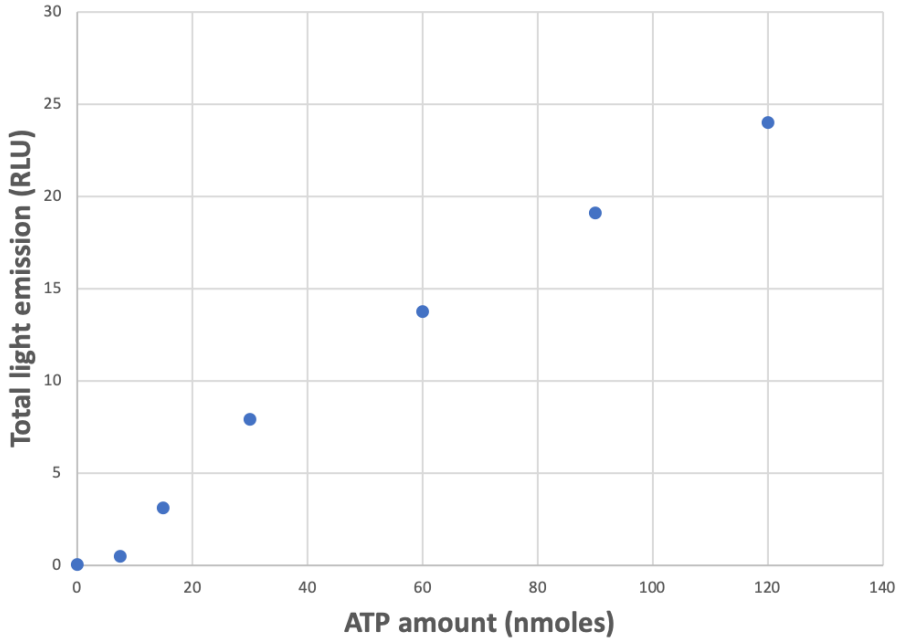


Figure 5.22– Calibration of ATP concentration with the Luc-mcRFP fibre. (The sum of the integrated light emission for the total ATP added.)

5.3 Summary

Functional fibrous materials were successfully produced from the six fusion proteins both with and without the polyQ tag. The formation of the protein fibres was achieved from the thin-film on the AWI. SEM pictures showed recognisably similar features between the film and the fibre. The fibrous structures on the film resulted from the compression of the water surface. Different morphology of the fibre on the film was discovered for mcRFP and mcRFP-Q77, which confirmed the structural guidance of the polyQ tag. Such difference resulted from the mechanical behaviours for mcRFP and mcRFP-Q77 fibres. Even though mcRFP-Q77 fibres possess higher tensile strength due to the H-bond formed by polyQ structures, mcRFP fibres experienced a long necking behaviour during stretching. This behaviour was accounted for by the bundled structure in mcRFP fibre. A colour change was noticed under optical microscopy during stretching which could be attributes to chromophore distortion. The extensibility of the traditionally size-limited fibrous materials formed by disease-related peptides was tested experimentally for the first time. These fibres of several centimetres in lengths exhibits similar cross-sectional structure with silk and similar mechanical properties with collagen.

Different methods were developed to quantify the fibre fluorescent and bioluminescent activity and were applied to different fibres according to different properties of the fibres. With the mcRFP and mcRFP-Q77 fibres with purely fluorescent activity were found to be soluble in 8 M urea and thus the urea-dissolved BCA assay was used to quantify the total number of molecules in the fibre. The fluorescent activity was found to increase with increasing number of protein molecules (or fibre volume) in the fibre. 74% fluorescent activity was retained in mcRFP-Q77 fibres, whereas only 18% was retained in mcRFP fibres. Combining with the fact that high β-sheet content was observed in the compressed films, a conformational helical to sheet transition of Q77 was confirmed, which protected the deformation of mcRFP. This intramolecular interaction of polyQ was further confirmed by the largely increased blue proportion of the chromophore and the decrease in the green proportion. This indicates a further prevention of the dehydration step due to the intramolecular H-bond formed by polyQ.

A different case was obtained with Luc-mcRFP and its polyQ tagged partner. The addition of the polyQ showed no effect on the protection of the fluorescent intensity, both from the

assembly profile in Chapter 4, and the activity measurements in this Chapter. Furthermore, Luc-mcRFP-Q77 has the smallest increase in the blue chromophore, and the smallest decrease in the red chromophore, contrary to mcRFP-Q77, indicating the buried Q77 structure in LRQ.

The reaction of Luc contained fibres was more complicated. With a comparison between the homogeneous and inhomogeneous BCA reaction, it was confirmed that the reaction rate was largely varied with protein solutions and protein fibres and thus it was expected that slower kinetics would be observed for Luc contained fibre catalysed reaction, which was confirmed with the bioluminescent profile. However, a linear relationship was observed between the fluorescent activity and bioluminescent activity in the Luc-mcRFP dual functional fibre, with Luc activity about 3 orders magnitude less than fluorescent activity. Even though the activity of Luc was found to be largely decreased in all Luc contained constructs, the Luc contained fibres showed increased activity with increased fibre volumes, allowing them for further sensing applications. Furthermore, for the Luc-mcRFP fibre, light emission reached a plateau, suggesting that product inhibition did not occur. It is further proved that with this plateaued emission, the Luc-mcRFP fibre demonstrated a possibility to achieve an ATP sensing capability with longer linear range.

Chapter 6 Conclusions and Future Work

6.1 Summary of Findings

6.1.1 Activities and Structures of Fusion Proteins

Two structurally distinctive functional proteins, mcRFP and Luc, and their recombination, Luc-mcRFP, with the polyglutamine tagged counterpart, were successfully expressed and purified from *E. coli*. All the three polyQ tagged counterparts result in a low yield of pure proteins from the soluble part of lysate, indicating the formation of inclusion bodies by polyQ intramolecular interactions. The recombinant Luc-mcRFP illustrates an enhanced fluorescent intensity by 25% with an increase in the proportion the blue intermediate. Together with the higher QY and decreased fluorescent lifetimes for all three components of the chromophore, the changes in the structure for mcRFP was reviewed. It shows the prevention of the deprotonation to form the additional N-acrylimine, possibly due to the rigidified chromophore by the surrounding amino acid after the formation of the blue intermediate for this conformation, which resulted in an increase in the QY. However, an increase in solvent accessibility shortened the lifetime of the fluorescent chromophores. The fusion of Q77 on both mcRFP and Luc-mcRFP decreased the fluorescent intensity. An increased proportion of blue intermediate and red chromophore was observed for mcRFP-Q77 with unchanged QY and decreased lifetime; whereas for Luc-mcRFP-Q77, both the proportions of the green and blue intermediates increases, with increased QY and decreased lifetime. This is explained as the β-can structure due to the intramolecular H-bonds formed by the polyQ tag, which halt the deprotonation to form the second N-acrylimine from the blue intermediate but facilitate the oxidation of the green intermediate towards the red chromophore for mcRFP-Q77. The spatial protection of Luc-mcRFP-Q77 of Q77 by Luc and mcRFP, prevents the formation of the H-bond and thus the oxidation of the green intermediate to the red chromophore.

Fusion also has negative impact on the luciferase activity, but the insertion of mcRFP between Luc and Q77 rescued partial activity loss. Both Luc-Q77 and Luc-mcRFP-Q77 have similar substrates binding kinetics, whereas a largely decreased ATP binding was observed for Luc-mcRFP. Combining with the FT-IR, CD spectroscopy and molecular structural modelling, it is explained as a largely associated interaction between the two functional proteins in Luc-mcRFP. The folding of Q77 in between Luc-mcRFP protects the two structures in Luc-mcRFP-Q77.

6.1.2 Aggregated Protein Structures

It is observed that mcRFP-Q77 self-assembled into fibrous structures under long time incubation at room temperature, which is only occurred if mcRFP is heated. A variation of pH changed the morphology of heated mcRFP aggregation, whereas at the start of mcRFP-Q77 fibrous aggregation, different shapes was observed at different pH. ThT was used to detect the intramolecular β-sheet formation in protein aggregates. It is observed that polyQ tagged proteins have more β-sheet content in their aggregates. The sheet content in Luc-mcRFP is unexpectedly high, which is explained as the unfolded structure in its monomer facilitating this aggregation.

6.1.3 Kinetics of Protein Interfacial Assembly

The six recombinant protein constructs assemble at the interface with different kinetics. The fusion of Q77 on mcRFP largely increase the assembly speed and ability at each concentration and each protein amount tested, which shows the directing role of Q77 towards the interface. In contrast, Luc and Luc-mcRFP assembles fast at AWI on themselves. Q77 tag decreased the assembly speed of Luc and Luc-mcRFP, indicating the assembly to the interface was determined by the molecular large functional partners, minimising the effect of Q77. Comparing the molecular footprint on the water interface and the area occupied by the protein at each cross-sectional view, it shows a largely compressed secondary structure for each protein construct. A missing of collapse was observed for all of the Q77 tagged structures (mcRFP-Q77, Luc-Q77 and Luc-mcRFP-Q77), which provides a possible insight of the intermolecular backbone-side chain H-bond formation by polyQ.

6.1.4 Functional Protein Fibres

The protein films formed on the interface were collected and observed under microscope. Wrinkled structures were obtained for compressed mcRFP and mcRPF-Q77 films. The FT-IR illustrate a densely packed β-sheet structure, especially the transition of the α-helical structure for mcRFP-Q77 in solution to the β-sheet structure. Fibres were pulled from these wrinkles, which retained the fluorescence under microscope. The cross-sectional structures of these fibres mimic that of silk. The mechanical behaviour, however, falls in the same range with collagen. The increased elongation strain of mcRFP is accounted by the less H-bond formed comparing to mcRFP-Q77. The quantification of fluorescent intensity of the fibres indicates a higher fluorescent retention for Q77 tagged mcRFP, but not on the Luc-mcRFP, corresponding to the findings of the interfacial kinetics, that the large functional proteins direct the assembly instead of the polyQ tag. The fibres illustrate absorption wavelength shifts for the blue and green intermediates, which indicate the higher energy barrier for the deprotonation process to occur during unfolding and refolding steps of the fibre formation, which may be due to a “twist” in the bond with the cis-trans isomeric transformation in the chromophore. In general, all protein functionalities are proportional to either the volume of the fibre, or the total protein molecules in the fibre.

6.1.5 Protein Fibre for ATP Sensing

A protein fibre for ATP sensing was achieved with Luc-mcRFP, which illustrates the potential of sensing ability by such a mechanically superior material. The total light emission increased proportionally with ATP amount with a longer linear range until the highest amount detected, which is 120 nmoles, down to 7.5 nmoles. The non-decaying light emission property of the fibre bioluminescent reaction demonstrate this sensing ability. A reusable fibrous material is achieved with ATP sensing and fluorescent dual functionalities.

6.2 Limitations and Recommended Future Works

The first obvious limitation for this work is the activity loss from the initial protein fusion and the interfacial assembly. Better DNA construct designs will be beneficial to maintain a higher activity for both proteins. Secondly, uniformly produced fibre materials from the air-water interface will be an advantage for the later activity test and sensing application. This may be achieved by carefully design of the protein structure either through de novo design or chemical modifications for a better controlled assembly structure. In addition, a mechanical rotor might be used to achieve a better controlled size and scaled-up production of these fibrous materials. Multiple LB troughs with smaller surface area, together with the mechanical rotor production method will allow batch production of these protein fibres. The uniformly produced fibres will consequently allow more controlled study of fibre properties such as substrate kinetics, stability, and ATP sensing. It will also eliminate the difficult step to quantify the total amount of proteins in each fibre. Under such circumstance, the molecular packing density in the fibre should be calculated as an important fibre characteristic. Many other basic studies for fibre productions such as fibre drying should be carried out. Finally, the detection range and sensitivity should be studied for the ATP sensing with LR fibre. Such sensing capability could also be tested with other enzymes to further confirm the general application of this platform.

6.3 Conclusions

In conclusion, a self-supportive functional protein fibrous product was developed from fusion protein interfacial assembly. The role of the polyglutamine tag was studied. The polyQ tag is necessary for the assembly of mcRFP, which has a β-barrel structure. The polyQ tag facilitate the molecular assembly at the interface. The resultant mcRFP-Q77 fibre also shows higher retained fluorescent activity and exhibits a higher mechanical strength. In comparison, firefly luciferase molecules and the Luc-mcRFP fused molecules assemble fast on the interface. The polyQ tag, however, lowers the assembly speed, and shows no effect on the protection of the functional activities. Therefore, it is concluded that the assembly-directed role of polyQ at AWI is dependent on its fusion partner. The dual functional Luc-mcRFP fibre was successfully used

in an ATP sensing assay. It demonstrates the potential use of such fibrous platform in sensing application.

References

1. Chothia C, Hubbard T, Brenner S, Barns H, Murzin A. PROTEIN FOLDS IN THE ALL-β AND ALL-α CLASSES. *Annu Rev Biophys Biomol Struct*. 1997;26(1):597-627. doi:10.1146/annurev.biophys.26.1.597

2. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and Combining Predictors of Mostly Disordered Proteins †. *Biochemistry*. 2005;44(6):1989-2000. doi:10.1021/bi047993o

3. Newberry RW, Raines RT. Secondary Forces in Protein Folding. *ACS Chem Biol*. 2019;14(8):1677-1686. doi:10.1021/acschembio.9b00339

4. Abaskharon RM, Gai F. Direct measurement of the tryptophan-mediated photocleavage kinetics of a protein disulfide bond. *Phys Chem Chem Phys*. 2016;18(14):9602-9607. doi:10.1039/C6CP00865H

5. Polypeptide chain configurations in crystalline proteins. *Proc R Soc London Ser A Math Phys Sci*. 1950;203(1074):321-357. doi:10.1098/rspa.1950.0142

6. Liu J, Zheng Q, Deng Y, Cheng C-S, Kallenbach NR, Lu M. A seven-helix coiled coil. *Proc Natl Acad Sci*. 2006;103(42):15457-15462. doi:10.1073/pnas.0604871103

7. Hakoshima T. Leucine Zippers. In: *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons, Ltd; 2005. doi:10.1038/npg.els.0005049

8. Du N, Liu XY, Narayanan J, Li L, Lim MLM, Li D. Design of Superior Spider Silk: From Nanostructure to Mechanical Properties. *Biophys J*. 2006;91(12):4528-4535. doi:10.1529/biophysj.106.089144

9. Wenzell NA, Ganguly HK, Pandey AK, Bhatt MR, Yap GPA, Zondlo NJ. Electronic and Steric Control of n→π* Interactions: Stabilization of the α-Helix Conformation without a Hydrogen Bond. *ChemBioChem*. 2019;20(7):963-967. doi:10.1002/cbic.201800785

10. Chiti F, Dobson CM. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu Rev Biochem*. 2006;75(1):333-366. doi:10.1146/annurev.biochem.75.101304.123901

11. Pauling L, Corey RB, Branson HR. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci*. 1951;37(4):205-211. doi:10.1073/pnas.37.4.205

12. Eisenberg D. The discovery of the -helix and -sheet, the principal structural features of proteins. *Proc Natl Acad Sci*. 2003;100(20):11207-11210. doi:10.1073/pnas.2034522100

13. Chothia C. Principles that Determine the Structure of Proteins. *Annu Rev Biochem*. 1984;53(1):537-572. doi:10.1146/annurev.bi.53.070184.002541

14. Nina Parker, Mark Schneegurt, Anh-Hue Thi Tu, Philip Lister BMF. https://openstax.org/books/microbiology/pages/7-4-proteins. In: *Microbiology*. Houston, Texas. https://openstax.org/books/microbiology/pages/1-introduction.

15. Samson AL, Ho B, Au AE, et al. Physicochemical properties that control protein aggregation also determine whether a protein is retained or released from necrotic cells. *Open Biol*. 2016;6(11):160098. doi:10.1098/rsob.160098

16. Carrell RW, Lomas DA. Conformational disease. *Lancet*. 1997;350(9071):134-138. doi:10.1016/S0140-6736(97)02073-4

17. Soto C. Protein misfolding and disease; protein refolding and therapy. *FEBS Lett*. 2001;498(2-3):204-207. doi:10.1016/S0014-5793(01)02486-3

18. Verma D, Gulati N, Kaul S, Mukherjee S, Nagaich U. Protein Based Nanostructures for Drug Delivery. *J Pharm*. 2018;2018:1-18. doi:10.1155/2018/9285854

19. Khadka DB, Haynie DT. Protein- and peptide-based electrospun nanofibers in medical biomaterials. *Nanomedicine Nanotechnology, Biol Med*. 2012;8(8):1242-1262. doi:10.1016/j.nano.2012.02.013

20. Decher G. Book Review: An Introduction to Ultrathin Organic Films from Langmuir-Blodgett to Self-Assembly. By A. Ulman. *Angew Chemie Int Ed English*. 1992;31(7):929-930. doi:10.1002/anie.199209291

21. Jiang H, Sang Y, Zhang L, Liu M. Self-Assembly and Directed Assembly. In: *Materials Nanoarchitectonics*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2018:165-186. doi:10.1002/9783527808311.ch9

22. Ranji A, Wu JC, Bundy BC, Jewett MC. Transforming Synthetic Biology with Cell-Free Systems. In: *Synthetic Biology*. Elsevier; 2013:277-301. doi:10.1016/B978-0-12-394430-6.00015-7

23. Zhang S, Marini DM, Hwang W, Santoso S. Design of nanostructured biological materials through self-assembly of peptides and proteins. *Curr Opin Chem Biol*. 2002;6(6):865-871. http://www.ncbi.nlm.nih.gov/pubmed/12470743.

24. Celikkin N, Rinoldi C, Costantini M, Trombetta M, Rainer A, Święszkowski W. Naturally derived proteins and glycosaminoglycan scaffolds for tissue engineering applications. *Mater Sci Eng C*. 2017;78:1277-1299. doi:10.1016/j.msec.2017.04.016

25. McManus JJ, Charbonneau P, Zaccarelli E, Asherie N. The physics of protein self-assembly. *Curr Opin Colloid Interface Sci*. 2016;22:73-79. doi:10.1016/j.cocis.2016.02.011

26. Ji X-T, Huang L, Huang H-Q. Construction of nanometer cisplatin core-ferritin (NCC-

F) and proteomic analysis of gastric cancer cell apoptosis induced with cisplatin released from the NCC-F. *J Proteomics*. 2012;75(11):3145-3157. doi:10.1016/j.jprot.2012.03.013

27. Ochs F, Karemore G, Miron E, et al. Stabilization of chromatin topology safeguards genome integrity. *Nature*. 2019;574(7779):571-574. doi:10.1038/s41586-019-1659-4

28. Fan C, Cheng S, Liu Y, et al. Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci*. 2010;107(16):7509-7514. doi:10.1073/pnas.0913199107

29. Shen C-H. Gene Expression: Translation of the Genetic Code. In: *Diagnostic Molecular Biology*. Elsevier; 2019:87-116. doi:10.1016/B978-0-12-802823-0.00004-3

30. Vepari C, Kaplan DL. Silk as a biomaterial. *Prog Polym Sci*. 2007;32(8-9):991-1007. doi:10.1016/j.progpolymsci.2007.05.013

31. Sasso L, Gerrard JA. Self-Assembled Biological Nanofibers for Biosensor Applications. In: *Micro and Nanofabrication Using Self-Assembled Biological Nanostructures*. Elsevier; 2015:1-20. doi:10.1016/B978-0-323-29642-7.00001-1

32. Hosseinkhani H, Hong P-D, Yu D-S. Self-Assembled Proteins and Peptides for Regenerative Medicine. *Chem Rev*. 2013;113(7):4837-4861. doi:10.1021/cr300131h

33. Nagamune T. Biomolecular engineering for nanobio/bionanotechnology. *Nano Converg*. 2017;4(1):9. doi:10.1186/s40580-017-0103-4

34. van der Linden E, Venema P. Self-assembly and aggregation of proteins. *Curr Opin Colloid Interface Sci*. 2007;12(4-5):158-165. doi:10.1016/j.cocis.2007.07.010

35. Fu Z, Luo Y, Derreumaux P, Wei G. Induced β-Barrel Formation of the Alzheimer's Aβ25–35 Oligomers on Carbon Nanotube Surfaces: Implication for Amyloid Fibril Inhibition. *Biophys J*. 2009;97(6):1795-1803. doi:10.1016/j.bpj.2009.07.014

36. Dumoulin M, Kumita JR, Dobson CM. Normal and Aberrant Biological Self-Assembly: Insights from Studies of Human Lysozyme and Its Amyloidogenic Variants. *Acc Chem Res*. 2006;39(9):603-610. doi:10.1021/ar050070g

37. Poirier MA, Li H, Macosko J, Cai S, Amzel M, Ross CA. Huntingtin Spheroids and Protofibrils as Precursors in Polyglutamine Fibrilization. *J Biol Chem*. 2002;277(43):41032-41037. doi:10.1074/jbc.M205809200

38. Domigan LJ, Healy JP, Meade SJ, Blaikie RJ, Gerrard JA. Controlling the dimensions of amyloid fibrils: Toward homogenous components for bionanotechnology. *Biopolymers*. 2012;97(2):123-133. doi:10.1002/bip.21709

39. Goers J, Permyakov SE, Permyakov EA, Uversky VN, Fink AL. Conformational Prerequisites for α-Lactalbumin Fibrillation †. *Biochemistry*. 2002;41(41):12546-12551. doi:10.1021/bi0262698

40. Graveland-Bikker JF, Ipsen R, Otte J, de Kruif CG. Influence of Calcium on the Self-Assembly of Partially Hydrolyzed α-Lactalbumin. *Langmuir*. 2004;20(16):6841-6846. doi:10.1021/la049579v

41. Fei L, Perrett S. Effect of Nanoparticles on Protein Folding and Fibrillogenesis. *Int J Mol Sci*. 2009;10(2):646-655. doi:10.3390/ijms10020646

42. Yang T, Zhang Y, Li Z. Formation of Gold Nanoparticle Decorated Lysozyme Microtubes. *Biomacromolecules*. 2011;12(6):2027-2031. doi:10.1021/bm2003835

43. Parveen R, Shamsi TN, Fatima S. Nanoparticles-protein interaction: Role in protein aggregation and clinical implications. *Int J Biol Macromol*. 2017;94:386-395. doi:10.1016/j.ijbiomac.2016.10.024

44. Schleeger M, VandenAkker CC, Deckert-Gaudig T, et al. Amyloids: From molecular structure to mechanical properties. *Polymer (Guildf)*. 2013;54(10):2473-2488. doi:10.1016/j.polymer.2013.02.029

45. Ridgley D, Rippner C, Barone J. Design and Construction of Large Amyloid Fibers. *Fibers*. 2015;3(4):90-102. doi:10.3390/fib3020090

46. Kammerer RA, Kostrewa D, Zurdo J, et al. Exploring amyloid formation by a de novo design. *Proc Natl Acad Sci*. 2004;101(13):4435-4440. doi:10.1073/pnas.0306786101

47. Potekhin S., Melnik T., Popov V, et al. De novo design of fibrils made of short α-helical coiled coil peptides. *Chem Biol*. 2001;8(11):1025-1032. doi:10.1016/S1074-5521(01)00073-4

48. Shen H, Fallas JA, Lynch E, et al. De novo design of self-assembling helical protein filaments. *Science (80-)*. 2018;362(6415):705-709. doi:10.1126/science.aau3775

49. Hammond JL, Formisano N, Estrela P, Carrara S, Tkac J. Electrochemical biosensors and nanobiosensors. Estrela P, ed. *Essays Biochem*. 2016;60(1):69-80. doi:10.1042/EBC20150008

50. Le Goff A, Holzinger M, Cosnier S. Enzymatic biosensors based on SWCNT-conducting polymer electrodes. *Analyst*. 2011;136(7):1279. doi:10.1039/c0an00904k

51. Anik Ü. Electrochemical medical biosensors for POC applications. In: *Medical Biosensors for Point of Care (POC) Applications*. Elsevier; 2017:275-292. doi:10.1016/B978-0-08-100072-4.00012-5

52. Quinchia J, Echeverri D, Cruz-Pacheco A, Maldonado M, Orozco J. Electrochemical Biosensors for Determination of Colorectal Tumor Biomarkers. *Micromachines*. 2020;11(4):411. doi:10.3390/mi11040411

53. Akolpoglu MB, Bozuyuk U, Erkoc P, Kizilel S. Biosensing–Drug Delivery Systems for In Vivo Applications. In: *Advanced Biosensors for Health Care Applications*. Elsevier; 2019:249-262. doi:10.1016/B978-0-12-815743-5.00009-3

54. Scheibel T, Parthasarathy R, Sawicki G, Lin X-M, Jaeger H, Lindquist SL. Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proc Natl Acad Sci*. 2003;100(8):4527-4532. doi:10.1073/pnas.0431081100

55. Wang J, Zhao X, Li J, et al. Electrostatic Assembly of Peptide Nanofiber–Biomimetic Silver Nanowires onto Graphene for Electrochemical Sensors. *ACS Macro Lett*.

2014;3(6):529-533. doi:10.1021/mz500213w

56. Carny O, Shalev DE, Gazit E. Fabrication of Coaxial Metal Nanocables Using a Self-Assembled Peptide Nanotube Scaffold. *Nano Lett*. 2006;6(8):1594-1597. doi:10.1021/nl060468l

57. Meier C, Lifincev I, Welland ME. Conducting Core–Shell Nanowires by Amyloid Nanofiber Templated Polymerization. *Biomacromolecules*. 2015;16(2):558-563. doi:10.1021/bm501618c

58. Hamedi M, Herland A, Karlsson RH, Inganäs O. Electrochemical Devices Made from Conducting Nanowire Networks Self-Assembled from Amyloid Fibrils and Alkoxysulfonate PEDOT. *Nano Lett*. 2008;8(6):1736-1740. doi:10.1021/nl0808233

59. Baldwin AJ, Bader R, Christodoulou J, MacPhee CE, Dobson CM, Barker PD. Cytochrome Display on Amyloid Fibrils. *J Am Chem Soc*. 2006;128(7):2162-2163. doi:10.1021/ja0565673

60. Adler-Abramovich L, Aronov D, Beker P, et al. Self-assembled arrays of peptide nanotubes by vapour deposition. *Nat Nanotechnol*. 2009;4(12):849-854. doi:10.1038/nnano.2009.298

61. Sopher NB, Abrams ZR, Reches M, Gazit E, Hanein Y. Integrating peptide nanotubes in micro-fabrication processes. *J Micromechanics Microengineering*. 2007;17(11):2360-2365. doi:10.1088/0960-1317/17/11/025

62. Castillo-León J, Rodriguez-Trujillo R, Gauthier S, Jensen ACØ, Svendsen WE. Micro-"factory" for self-assembled peptide nanostructures. *Microelectron Eng*. 2011;88(8):1685-1688. doi:10.1016/j.mee.2010.12.023

63. Yan X, Zhu P, Li J. Self-assembly and application of diphenylalanine-based nanostructures. *Chem Soc Rev*. 2010;39(6):1877. doi:10.1039/b915765b

64. Ivnitski D, Amit M, Silberbush O, et al. The Strong Influence of Structure Polymorphism on the Conductivity of Peptide Fibrils. *Angew Chemie Int Ed*. 2016;55(34):9988-9992. doi:10.1002/anie.201604833

65. Bal S, Ghosh C, Ghosh T, Vijayaraghavan RK, Das D. Non-Equilibrium Polymerization of Cross-β Amyloid Peptides for Temporal Control of Electronic Properties. *Angew Chemie Int Ed*. May 2020:anie.202003721. doi:10.1002/anie.202003721

66. Amit M, Yuran S, Gazit E, Reches M, Ashkenasy N. Tailor-Made Functional Peptide Self-Assembling Nanostructures. *Adv Mater*. 2018;30(41):1707083. doi:10.1002/adma.201707083

67. Zhuravel R, Amit E, Elbaz S, et al. Atomic force microscopy characterization of kinase-mediated phosphorylation of a peptide monolayer. *Sci Rep*. 2016;6(1):36793. doi:10.1038/srep36793

68. Kim JH, Lee M, Lee JS, Park CB. Self-Assembled Light-Harvesting Peptide Nanotubes for Mimicking Natural Photosynthesis. *Angew Chemie Int Ed*.

2012;51(2):517-520. doi:10.1002/anie.201103244

69. Park J, Lee W, Lee G, et al. Electrochemical Amyloid-Based Biosensor for the Determination of Metal Ions. *J Electrochem Soc*. 2019;166(15):B1497-B1505. doi:10.1149/2.0451915jes

70. Li C, Adamcik J, Mezzenga R. Biodegradable nanocomposites of amyloid fibrils and graphene with shape-memory and enzyme-sensing properties. *Nat Nanotechnol*. 2012;7(7):421-427. doi:10.1038/nnano.2012.62

71. Wang L, Chen D, Jiang K, Shen G. New insights and perspectives into biological materials for flexible electronics. *Chem Soc Rev*. 2017;46(22):6764-6815. doi:10.1039/C7CS00278E

72. Baxa U, Speransky V, Steven AC, Wickner RB. Mechanism of inactivation on prion conversion of the Saccharomyces cerevisiae Ure2 protein. *Proc Natl Acad Sci*. 2002;99(8):5253-5260. doi:10.1073/pnas.082097899

73. Totzeck F, Andrade-Navarro MA, Mier P. The Protein Structure Context of PolyQ Regions. van der Wel P, ed. *PLoS One*. 2017;12(1):e0170801. doi:10.1371/journal.pone.0170801

74. Lajoie P, Snapp EL. Formation and Toxicity of Soluble Polyglutamine Oligomers in Living Cells. Chirico G, ed. *PLoS One*. 2010;5(12):e15245. doi:10.1371/journal.pone.0015245

75. Leitman J, Ulrich Hartl F, Lederkremer GZ. Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. *Nat Commun*. 2013;4(1):2753. doi:10.1038/ncomms3753

76. Lu M, Banetta L, Young LJ, et al. Live-cell super-resolution microscopy reveals a primary role for diffusion in polyglutamine-driven aggresome assembly. *J Biol Chem*. 2019;294(1):257-268. doi:10.1074/jbc.RA118.003500

77. Chow MKM, Paulson HL, Bottomley SP. Destabilization of a Non-pathological Variant of Ataxin-3 Results in Fibrillogenesis via a Partially Folded Intermediate: A Model for Misfolding in Polyglutamine Disease. *J Mol Biol*. 2004;335(1):333-341. doi:10.1016/j.jmb.2003.08.064

78. Shehi E, Fusi P, Secundo F, Pozzuolo S, Bairati A, Tortora P. Temperature-Dependent, Irreversible Formation of Amyloid Fibrils by a Soluble Human Ataxin-3 Carrying a Moderately Expanded Polyglutamine Stretch (Q36) †. *Biochemistry*. 2003;42(49):14626-14632. doi:10.1021/bi0352825

79. Marchal S, Shehi E, Harricane M-C, et al. Structural Instability and Fibrillar Aggregation of Non-expanded Human Ataxin-3 Revealed under High Pressure and Temperature. *J Biol Chem*. 2003;278(34):31554-31563. doi:10.1074/jbc.M304205200

80. Streets AM, Sourigues Y, Kopito RR, Melki R, Quake SR. Simultaneous Measurement of Amyloid Fibril Formation by Dynamic Light Scattering and Fluorescence Reveals Complex Aggregation Kinetics. Georgakoudi I, ed. *PLoS One*. 2013;8(1):e54541. doi:10.1371/journal.pone.0054541

81. Chow MKM, Ellisdon AM, Cabrita LD, Bottomley SP. Polyglutamine Expansion in Ataxin-3 Does Not Affect Protein Stability. *J Biol Chem*. 2004;279(46):47643-47651. doi:10.1074/jbc.M405799200

82. Scherzinger E, Lurz R, Turmaine M, et al. Huntingtin-Encoded Polyglutamine Expansions Form Amyloid-like Protein Aggregates In Vitro and In Vivo. *Cell*. 1997;90(3):549-558. doi:10.1016/S0092-8674(00)80514-0

83. Jiang Y, Di Gregorio SE, Duennwald ML, Lajoie P. Polyglutamine toxicity in yeast uncovers phenotypic variations between different fluorescent protein fusions. *Traffic*. 2017;18(1):58-70. doi:10.1111/tra.12453

84. Middelberg AP. Preparative protein refolding. *Trends Biotechnol*. 2002;20(10):437-443. doi:10.1016/S0167-7799(02)02047-4

85. Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. *Trends Biotechnol*. 2006;24(4):179-185. doi:10.1016/j.tibtech.2006.02.007

86. Ami D, Natalello A, Taylor G, Tonon G, Maria Doglia S. Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. *Biochim Biophys Acta - Proteins Proteomics*. 2006;1764(4):793-799. doi:10.1016/j.bbapap.2005.12.005

87. Rhys NH, Dougan L. The emerging role of hydrogen bond interactions in polyglutamine structure, stability and association. *Soft Matter*. 2013;9(8):2359-2364. doi:10.1039/C2SM27565A

88. Shimomura O, Johnson FH, Saiga Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *J Cell Comp Physiol*. 1962;59(3):223-239. doi:10.1002/jcp.1030590302

89. Matz M V., Fradkov AF, Labas YA, et al. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol*. 1999;17(10):969-973. doi:10.1038/13657

90. Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci*. 2000;97(22):11990-11995. doi:10.1073/pnas.97.22.11990

91. Strack RL, Strongin DE, Mets L, Glick BS, Keenan RJ. Chromophore Formation in DsRed Occurs by a Branched Pathway. *J Am Chem Soc*. 2010;132(24):8496-8505. doi:10.1021/ja1030084

92. Subach F V., Verkhusha V V. Chromophore Transformations in Red Fluorescent Proteins. *Chem Rev*. 2012;112(7):4308-4327. doi:10.1021/cr2001965

93. Campbell RE, Tour O, Palmer AE, et al. A monomeric red fluorescent protein. *Proc Natl Acad Sci*. 2002;99(12):7877-7882. doi:10.1073/pnas.082243699

94. Branchini BR, Southworth TL, Fontaine DM, Kohrt D, Florentine CM, Grossel MJ. A Firefly Luciferase Dual Color Bioluminescence Reporter Assay Using Two Substrates To Simultaneously Monitor Two Gene Expression Events. *Sci Rep*. 2018;8(1):5990. doi:10.1038/s41598-018-24278-2

95. Kirkpatrick A, Xu T, Ripp S, Sayler G, Close D. Biotechnological Advances in Luciferase Enzymes. In: *Bioluminescence - Analytical Applications and Basic Biology*. IntechOpen; 2019. doi:10.5772/intechopen.85313

96. Widder EA, Falls B. Review of Bioluminescence for Engineers and Scientists in Biophotonics. *IEEE J Sel Top Quantum Electron*. 2014;20(2):232-241. doi:10.1109/JSTQE.2013.2284434

97. Baldwin TO. Firefly luciferase: the structure is known, but the mystery remains. *Structure*. 1996;4(3):223-228. doi:10.1016/S0969-2126(96)00026-3

98. White EH, Rapaport E, Hopkins TA, Seliger HH. Chemi- and bioluminescence of firefly luciferin. *J Am Chem Soc*. 1969;91(8):2178-2180. doi:10.1021/ja01036a093

99. Nakatani N, Hasegawa J, Nakatsuji H. Red Light in Chemiluminescence and Yellow-Green Light in Bioluminescence: Color-Tuning Mechanism of Firefly, Photinus pyralis , Studied by the Symmetry-Adapted Cluster–Configuration Interaction Method. *J Am Chem Soc*. 2007;129(28):8756-8765. doi:10.1021/ja0611691

100. Seliger HH, McElroy WD. THE COLORS OF FIREFLY BIOLUMINESCENCE: ENZYME CONFIGURATION AND SPECIES SPECIFICITY. *Proc Natl Acad Sci*. 1964;52(1):75-81. doi:10.1073/pnas.52.1.75

101. Law GHE, Gandelman OA, Tisi LC, Lowe CR, Murray JAH. Mutagenesis of solvent-exposed amino acids in Photinus pyralis luciferase improves thermostability and pH-tolerance. *Biochem J*. 2006;397(2):305-312. doi:10.1042/BJ20051847

102. Branchini BR, Southworth TL, Khattak NF, Michelini E, Roda A. Red- and green-emitting firefly luciferase mutants for bioluminescent reporter applications. *Anal Biochem*. 2005;345(1):140-148. doi:10.1016/j.ab.2005.07.015

103. Gabriel GV de M, Yasuno R, Mitani Y, Ohmiya Y, Viviani VR. Novel application of Macrolampis sp2 firefly luciferase for intracellular pH-biosensing in mammalian cells. *Photochem Photobiol Sci*. 2019;18(5):1212-1217. doi:10.1039/C8PP00573G

104. Borghei G, Hall EAH. BRET-linked ATP assay with luciferase. *Analyst*. 2014;139(17):4185-4192. doi:10.1039/C4AN00436A

105. Andreeva A, Howorth D, Chothia C, Kulesha E, Murzin AG. SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Res*. 2014;42(D1):D310-D314. doi:10.1093/nar/gkt1242

106. Di Lullo GA, Sweeney SM, Körkkö J, Ala-Kokko L, San Antonio JD. Mapping the Ligand-binding Sites and Disease-associated Mutations on the Most Abundant Protein in the Human, Type I Collagen. *J Biol Chem*. 2002;277(6):4223-4231. doi:10.1074/jbc.M110709200

107. Ricard-Blum S. The Collagen Family. *Cold Spring Harb Perspect Biol*. 2011;3(1):a004978-a004978. doi:10.1101/cshperspect.a004978

108. Squire J. Special Issue: The Actin-Myosin Interaction in Muscle: Background and Overview. *Int J Mol Sci*. 2019;20(22):5715. doi:10.3390/ijms20225715

109. Rouse JG, Van Dyke ME. A Review of Keratin-Based Biomaterials for Biomedical Applications. *Materials (Basel)*. 2010;3(2):999-1014. doi:10.3390/ma3020999

110. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Collagen: The Fibrous Proteins of the Matrix. In: *Molecular Cell Biology*. 4th ed. ; 2000. https://www.ncbi.nlm.nih.gov/books/NBK21582/.

111. Kozel BA, Mecham RP. Elastic fiber ultrastructure and assembly. *Matrix Biol*. 2019;84:31-40. doi:10.1016/j.matbio.2019.10.002

112. Jensen SA, Reinhardt DP, Gibson MA, Weiss AS. Protein Interaction Studies of MAGP-1 with Tropoelastin and Fibrillin-1. *J Biol Chem*. 2001;276(43):39661-39666. doi:10.1074/jbc.M104533200

113. Roark EF, Keene DR, Haudenschild CC, Godyna S, Little CD, Argraves WS. The association of human fibulin-1 with elastic fibers: an immunohistological, ultrastructural, and RNA study. *J Histochem Cytochem*. 1995;43(4):401-411. doi:10.1177/43.4.7534784

114. Li B, Daggett V. Molecular basis for the extensibility of elastin. *J Muscle Res Cell Motil*. 2002;23:561–573. doi:https://doi.org/10.1023/A:1023474909980

115. DAAMEN W, VEERKAMP J, VANHEST J, VANKUPPEVELT T. Elastin as a biomaterial for tissue engineering. *Biomaterials*. 2007;28(30):4378-4398. doi:10.1016/j.biomaterials.2007.06.025

116. Yassine NM, Shahram JT, Body SC. Pathogenic Mechanisms of Bicuspid Aortic Valve Aortopathy. *Front Physiol*. 2017;8. doi:10.3389/fphys.2017.00687

117. Daniels F, ter Haar Romeny BM, Rubbens M, van Assen H. Quantification of Collagen Orientation in 3D Engineered Tissue. In: ; 2007:282-286. doi:10.1007/978-3-540-68017-8_73

118. Feughelman M. Natural protein fibers. *J Appl Polym Sci*. 2002;83(3):489-507. doi:10.1002/app.2255

119. Cai T, Han K, Yang P, et al. Reconstruction of Dynamic and Reversible Color Change using Reflectin Protein. *Sci Rep*. 2019;9(1):5201. doi:10.1038/s41598-019-41638-8

120. Guan Z, Cai T, Liu Z, et al. Origin of the Reflectin Gene and Hierarchical Assembly of Its Protein. *Curr Biol*. 2017;27(18):2833-2842.e6. doi:10.1016/j.cub.2017.07.061

121. Xu C, Stiubianu GT, Gorodetsky AA. Adaptive infrared-reflecting systems inspired by cephalopods. *Science (80-)*. 2018;359(6383):1495-1500. doi:10.1126/science.aar5191

122. Chaudhary G, Ewoldt RH, Thiffeault J-L. Unravelling hagfish slime. *J R Soc Interface*. 2019;16(150):20180710. doi:10.1098/rsif.2018.0710

123. Koch EA, Spitzer RH, Pithawalla RB, Castillos FA, Parry DAD. Hagfish biopolymer: a type I/type II homologue of epidermal keratin intermediate filaments. *Int J Biol Macromol*. 1995;17(5):283-292. doi:10.1016/0141-8130(95)98156-S

124. Koch EA, Spitzer RH, Pithawalla RB, Downing SW. Keratin-like components of

gland thread cells modulate the properties of mucus from hagfish (Eptatretus stouti). *Cell Tissue Res*. 1991;264(1):79-86. doi:10.1007/BF00305724

125. Fudge DS. Composition, morphology and mechanics of hagfish slime. *J Exp Biol*. 2005;208(24):4613-4625. doi:10.1242/jeb.01963

126. Santin M, Motta A, Freddi G, Cannas M. In vitro evaluation of the inflammatory potential of the silk fibroin. *J Biomed Mater Res*. 1999;46(3):382-389. doi:10.1002/(SICI)1097-4636(19990905)46:3<382::AID-JBM11>3.0.CO;2-R

127. Acharya C, Kumar V, Sen R, Kundu SC. Performance evaluation of a silk protein-based matrix for the enzymatic conversion of tyrosine toL-DOPA. *Biotechnol J*. 2008;3(2):226-233. doi:10.1002/biot.200700120

128. Unger RE, Peters K, Wolf M, Motta A, Migliaresi C, Kirkpatrick CJ. Endothelialization of a non-woven silk fibroin net for use in tissue engineering: growth and gene regulation of human endothelial cells. *Biomaterials*. 2004;25(21):5137-5146. doi:10.1016/j.biomaterials.2003.12.040

129. Dal Pra I, Freddi G, Minic J, Chiarini A, Armato U. De novo engineering of reticular connective tissue in vivo by silk fibroin nonwoven materials. *Biomaterials*. 2005;26(14):1987-1999. doi:10.1016/j.biomaterials.2004.06.036

130. Barua E, Deoghare AB, Deb P, Lala S Das. Naturally derived biomaterials for development of composite bone scaffold: A review. *IOP Conf Ser Mater Sci Eng*. 2018;377:012013. doi:10.1088/1757-899X/377/1/012013

131. Acharya C, Ghosh SK, Kundu SC. Silk fibroin protein from mulberry and non-mulberry silkworms: cytotoxicity, biocompatibility and kinetics of L929 murine fibroblast adhesion. *J Mater Sci Mater Med*. 2008;19(8):2827-2836. doi:10.1007/s10856-008-3408-3

132. Kundu J, Dewan M, Ghoshal S, Kundu SC. Mulberry non-engineered silk gland protein vis-à-vis silk cocoon protein engineered by silkworms as biomaterial matrices. *J Mater Sci Mater Med*. 2008;19(7):2679-2689. doi:10.1007/s10856-008-3398-1

133. Inoue S, Tanaka K, Arisaka F, Kimura S, Ohtomo K, Mizuno S. Silk Fibroin of Bombyx mori Is Secreted, Assembling a High Molecular Mass Elementary Unit Consisting of H-chain, L-chain, and P25, with a 6:6:1 Molar Ratio. *J Biol Chem*. 2000;275(51):40517-40528. doi:10.1074/jbc.M006897200

134. Altman GH, Diaz F, Jakuba C, et al. Silk-based biomaterials. *Biomaterials*. 2003;24(3):401-416. doi:10.1016/S0142-9612(02)00353-8

135. Mondal M, Trivedy K, Kumar N. The silk proteins, sericin and fibroin in silkworm, Bombyx mori Linn. - A review. 2006.

136. Jao D, Mou X, Hu X. Tissue Regeneration: A Silk Road. *J Funct Biomater*. 2016;7(3):22. doi:10.3390/jfb7030022

137. Blackledge TA, Kuntner M, Agnarsson I. The Form and Function of Spider Orb Webs. In: ; 2011:175-262. doi:10.1016/B978-0-12-415919-8.00004-5

138. Xu L, Rainey JK, Meng Q, Liu X-Q. Recombinant Minimalist Spider Wrapping Silk Proteins Capable of Native-Like Fiber Formation. Mitraki A, ed. *PLoS One*. 2012;7(11):e50227. doi:10.1371/journal.pone.0050227

139. Ling S, Qin Z, Li C, Huang W, Kaplan DL, Buehler MJ. Polymorphic regenerated silk fibers assembled through bioinspired spinning. *Nat Commun*. 2017;8(1):1387. doi:10.1038/s41467-017-00613-5

140. Weisman S, Haritos VS, Church JS, et al. Honeybee silk: Recombinant protein production, assembly and fiber spinning. *Biomaterials*. 2010;31(9):2695-2700. doi:10.1016/j.biomaterials.2009.12.021

141. Sutherland TD, Weisman S, Trueman HE, Sriskantha A, Trueman JWH, Haritos VS. Conservation of Essential Design Features in Coiled Coil Silks. *Mol Biol Evol*. 2007;24(11):2424-2432. doi:10.1093/molbev/msm171

142. Sutherland TD, Church JS, Hu X, Huson MG, Kaplan DL, Weisman S. Single Honeybee Silk Protein Mimics Properties of Multi-Protein Silk. Buehler M, ed. *PLoS One*. 2011;6(2):e16489. doi:10.1371/journal.pone.0016489

143. Rapson TD, Sutherland TD, Church JS, Trueman HE, Dacres H, Trowell SC. De Novo Engineering of Solid-State Metalloproteins Using Recombinant Coiled-Coil Silk. *ACS Biomater Sci Eng*. 2015;1(11):1114-1120. doi:10.1021/acsbiomaterials.5b00239

144. Ravichandran R, Islam MM, Alarcon EI, et al. Functionalised type-I collagen as a hydrogel building block for bio-orthogonal tissue engineering applications. *J Mater Chem B*. 2016;4(2):318-326. doi:10.1039/C5TB02035B

145. Ravichandran R, Astrand C, Patra HK, Turner APF, Chotteau V, Phopase J. Intelligent ECM mimetic injectable scaffolds based on functional collagen building blocks for tissue engineering and biomedical applications. *RSC Adv*. 2017;7(34):21068-21078. doi:10.1039/C7RA02927F

146. Vivekananthan V, Alluri NR, Purusothaman Y, Chandrasekhar A, Selvarajan S, Kim S-J. Biocompatible Collagen Nanofibrils: An Approach for Sustainable Energy Harvesting and Battery-Free Humidity Sensor Applications. *ACS Appl Mater Interfaces*. 2018;10(22):18650-18656. doi:10.1021/acsami.8b02915

147. Zhou Z, Qian D, Minary-Jolandan M. Molecular Mechanism of Polarization and Piezoelectric Effect in Super-Twisted Collagen. *ACS Biomater Sci Eng*. 2016;2(6):929-936. doi:10.1021/acsbiomaterials.6b00021

148. Jacob J, More N, Kalia K, Kapusetti G. Piezoelectric smart biomaterials for bone and cartilage tissue engineering. *Inflamm Regen*. 2018;38(1):2. doi:10.1186/s41232-018-0059-8

149. Hudspeth MA, Kaya T. Collagen as a Humidity Sensing Dielectric Material. *MRS Proc*. 2012;1427:mrss12-1427-b07-04. doi:10.1557/opl.2012.1415

150. Shapardanis S, Hudspeth M, Kaya T. Design and implementation of collagen-based capacitive relative humidity sensors. In: *2013 IEEE SENSORS*. IEEE; 2013:1-4. doi:10.1109/ICSENS.2013.6688506

151. Ravichandran R, Martinez JG, Jager EWH, Phopase J, Turner APF. Type I Collagen-Derived Injectable Conductive Hydrogel Scaffolds as Glucose Sensors. *ACS Appl Mater Interfaces*. 2018;10(19):16244-16249. doi:10.1021/acsami.8b04091

152. Wang X, Yue O, Liu X, Hou M, Zheng M. A novel bio-inspired multi-functional collagen aggregate based flexible sensor with multi-layer and internal 3D network structure. *Chem Eng J*. 2020;392:123672. doi:10.1016/j.cej.2019.123672

153. Fan S, Zhang Y, Huang X, et al. Silk materials for medical, electronic and optical applications. *Sci China Technol Sci*. 2019;62(6):903-918. doi:10.1007/s11431-018-9403-8

154. Wang X, Yucel T, Lu Q, Hu X, Kaplan DL. Silk nanospheres and microspheres from silk/pva blend films for drug delivery. *Biomaterials*. 2010;31(6):1025-1035. doi:10.1016/j.biomaterials.2009.11.002

155. Wongpinyochit T, Johnston BF, Seib FP. Manufacture and Drug Delivery Applications of Silk Nanoparticles. *J Vis Exp*. 2016;(116). doi:10.3791/54669

156. Hu D, Xu Z, Hu Z, Hu B, Yang M, Zhu L. pH-Triggered Charge-Reversal Silk Sericin-Based Nanoparticles for Enhanced Cellular Uptake and Doxorubicin Delivery. *ACS Sustain Chem Eng*. 2017;5(2):1638-1647. doi:10.1021/acssuschemeng.6b02392

157. Liu Z, Zhang M, Zhang Y, et al. Spider silk-based humidity sensor. *Opt Lett*. 2019;44(11):2907. doi:10.1364/OL.44.002907

158. Liu Z, Liu W, Hu C, et al. Natural spider silk as a photonics component for humidity sensing. *Opt Express*. 2019;27(15):21946. doi:10.1364/OE.27.021946

159. Kim HS, Cha SH, Roy B, Kim S, Ahn YH. Humidity sensing using THz metamaterial with silk protein fibroin. *Opt Express*. 2018;26(26):33575. doi:10.1364/OE.26.033575

160. Lee M, Jeon H, Kim S. A Highly Tunable and Fully Biocompatible Silk Nanoplasmonic Optical Sensor. *Nano Lett*. 2015;15(5):3358-3363. doi:10.1021/acs.nanolett.5b00680

161. Li X, Zong L, Wu X, You J, Li M, Li C. Biomimetic engineering of spider silk fibres with graphene for electric devices with humidity and motion sensitivity. *J Mater Chem C*. 2018;6(13):3212-3219. doi:10.1039/C8TC00265G

162. Wang C, Xia K, Zhang M, Jian M, Zhang Y. An All-Silk-Derived Dual-Mode E-skin for Simultaneous Temperature–Pressure Detection. *ACS Appl Mater Interfaces*. 2017;9(45):39484-39492. doi:10.1021/acsami.7b13356

163. Xu M, Yadavalli VK. Flexible Biosensors for the Impedimetric Detection of Protein Targets Using Silk-Conductive Polymer Biocomposites. *ACS Sensors*. 2019;4(4):1040-1047. doi:10.1021/acssensors.9b00230

164. Singh A, Hede S, Sastry M. Spider Silk as an Active Scaffold in the Assembly of Gold Nanoparticles and Application of the Gold–Silk Bioconjugate in Vapor Sensing. *Small*. 2007;3(3):466-473. doi:10.1002/smll.200600413

165. Wang Q, Wang C, Zhang M, Jian M, Zhang Y. Feeding Single-Walled Carbon

Nanotubes or Graphene to Silkworms for Reinforced Silk Fibers. *Nano Lett*. 2016;16(10):6695-6700. doi:10.1021/acs.nanolett.6b03597

166. Lee OJ, Sultan MT, Hong H, et al. Recent Advances in Fluorescent Silk Fibroin. *Front Mater*. 2020;7. doi:10.3389/fmats.2020.00050

167. Kumar M, Sanford KJ, Cuevas WA, Du M, Collier KD, Chow N. Designer Protein-Based Performance Materials. *Biomacromolecules*. 2006;7(9):2543-2551. doi:10.1021/bm060464a

168. Bini E, Foo CWP, Huang J, Karageorgiou V, Kitchel B, Kaplan DL. RGD-Functionalized Bioengineered Spider Dragline Silk Biomaterial. *Biomacromolecules*. 2006;7(11):3139-3145. doi:10.1021/bm0607877

169. Martín-Moldes Z, Ebrahimi D, Plowright R, et al. Intracellular Pathways Involved in Bone Regeneration Triggered by Recombinant Silk-Silica Chimeras. *Adv Funct Mater*. 2018;28(27):1702570. doi:10.1002/adfm.201702570

170. Horak J, Jansson R, Dev A, et al. Recombinant Spider Silk as Mediator for One-Step, Chemical-Free Surface Biofunctionalization. *Adv Funct Mater*. 2018;28(21):1800206. doi:10.1002/adfm.201800206

171. Wang Y, Xu S, Wang R, et al. Genetic fabrication of functional silk mats with improved cell proliferation activity for medical applications. *Biomater Sci*. 2019;7(11):4536-4546. doi:10.1039/C9BM01285K

172. Wang Y, Wang F, Xu S, et al. Genetically engineered bi-functional silk material with improved cell proliferation and anti-inflammatory activity for medical application. *Acta Biomater*. 2019;86:148-157. doi:10.1016/j.actbio.2018.12.036

173. Huang J, Wong C, George A, Kaplan DL. The effect of genetically engineered spider silk-dentin matrix protein 1 chimeric protein on hydroxyapatite nucleation. *Biomaterials*. 2007;28(14):2358-2367. doi:10.1016/j.biomaterials.2006.11.021

174. Numata K, Hamasaki J, Subramanian B, Kaplan DL. Gene delivery mediated by recombinant silk proteins containing cationic and cell binding motifs. *J Control Release*. 2010;146(1):136-143. doi:10.1016/j.jconrel.2010.05.006

175. Brenner MD, Zhou R, Conway DE, et al. Spider Silk Peptide Is a Compact, Linear Nanospring Ideal for Intracellular Tension Sensing. *Nano Lett*. 2016;16(3):2096-2102. doi:10.1021/acs.nanolett.6b00305

176. Jansson R, Courtin CM, Sandgren M, Hedhammar M. Rational Design of Spider Silk Materials Genetically Fused with an Enzyme. *Adv Funct Mater*. 2015;25(33):5343-5352. doi:10.1002/adfm.201501833

177. Ji W, Yang F, van den Beucken JJJP, et al. Fibrous scaffolds loaded with protein prepared by blend or coaxial electrospinning. *Acta Biomater*. 2010;6(11):4199-4207. doi:10.1016/j.actbio.2010.05.025

178. Wen D-L, Liu X, Deng H-T, et al. Printed silk-fibroin-based triboelectric nanogenerators for multi-functional wearable sensing. *Nano Energy*. 2019;66:104123.

doi:10.1016/j.nanoen.2019.104123

179. Porto MDA, dos Santos JP, Hackbart H, et al. Immobilization of α-amylase in ultrafine polyvinyl alcohol (PVA) fibers via electrospinning and their stability on different substrates. *Int J Biol Macromol*. 2019;126:834-841. doi:10.1016/j.ijbiomac.2018.12.263

180. Jiang W, Pei R, Zhou S-F. 3D-printed xylanase within biocompatible polymers as excellent catalyst for lignocellulose degradation. *Chem Eng J*. 2020;400:125920. doi:10.1016/j.cej.2020.125920

181. Parhi P, Golas A, Barnthip N, Noh H, Vogler EA. Volumetric interpretation of protein adsorption: Capacity scaling with adsorbate molecular weight and adsorbent surface energy. *Biomaterials*. 2009;30(36):6814-6824. doi:10.1016/j.biomaterials.2009.09.005

182. Krishnan A, Siedlecki CA, Vogler EA. Traube-Rule Interpretation of Protein Adsorption at the Liquid–Vapor Interface †. *Langmuir*. 2003;19(24):10342-10352. doi:10.1021/la035308t

183. Valluzzi R, Gido SP. The crystal structure ofBombyx mori silk fibroin at the air–water interface. *Biopolymers*. 1997;42(6):705-717. doi:10.1002/(SICI)1097-0282(199711)42:6<705::AID-BIP8>3.0.CO;2-Y

184. Morinaga A, Hasegawa K, Nomura R, et al. Critical role of interfaces and agitation on the nucleation of Aβ amyloid fibrils at low concentrations of Aβ monomers. *Biochim Biophys Acta - Proteins Proteomics*. 2010;1804(4):986-995. doi:10.1016/j.bbapap.2010.01.012

185. Campioni S, Carret G, Jordens S, Nicoud L, Mezzenga R, Riek R. The Presence of an Air–Water Interface Affects Formation and Elongation of α-Synuclein Fibrils. *J Am Chem Soc*. 2014;136(7):2866-2875. doi:10.1021/ja412105t

186. Jean L, Lee CF, Vaux DJ. Enrichment of Amyloidogenesis at an Air-Water Interface. *Biophys J*. 2012;102(5):1154-1162. doi:10.1016/j.bpj.2012.01.041

187. Chang C-H, Franses EI. Adsorption dynamics of surfactants at the air/water interface: a critical review of mathematical models, data, and mechanisms. *Colloids Surfaces A Physicochem Eng Asp*. 1995;100:1-45. doi:10.1016/0927-7757(94)03061-4

188. Singhal R, Chaubey A, Kaneto K, Takashima W, Malhotra BD. Poly-3-hexyl thiophene Langmuir-Blodgett films for application to glucose biosensor. *Biotechnol Bioeng*. 2004;85(3):277-282. doi:10.1002/bit.10869

189. Sharma SK, Singhal R, Malhotra BD, Sehgal N, Kumar A. Langmuir–Blodgett film based biosensor for estimation of galactose in milk. *Electrochim Acta*. 2004;49(15):2479-2485. doi:10.1016/j.electacta.2004.01.024

190. Sandt JD, Moudio M, Clark JK, et al. Stretchable Optomechanical Fiber Sensors for Pressure Determination in Compressive Medical Textiles. *Adv Healthc Mater*. 2018;7(15):1800293. doi:10.1002/adhm.201800293

191. Huang Z, Salim T, Brawley A, Patterson J, Matthews KS, Bondos SE.

Functionalization and Patterning of Protein-Based Materials Using Active Ultrabithorax Chimeras. *Adv Funct Mater*. 2011;21(14):2633-2640. doi:10.1002/adfm.201100067

192. Tsai S-P, Howell DW, Huang Z, et al. The Effect of Protein Fusions on the Production and Mechanical Properties of Protein-Based Materials. *Adv Funct Mater*. 2015;25(9):1442-1450. doi:10.1002/adfm.201402997

193. de Jongh HHJ, Kosters HA, Kudryashova E, Meinders MBJ, Trofimova D, Wierenga PA. Protein adsorption at air-water interfaces: A combination of details. *Biopolymers*. 2004;74(1-2):131-135. doi:10.1002/bip.20036

194. MacRitchie F. Protein adsorption/desorption at fluid interfaces. *Colloids and Surfaces*. 1989;41:25-34. doi:10.1016/0166-6622(89)80038-1

195. Ward AJI, Regan LH. Pendant drop studies of adsorbed films of bovine serum albumin. *J Colloid Interface Sci*. 1980;78(2):389-394. doi:10.1016/0021-9797(80)90578-0

196. Tornberg E. The application of the drop volume technique to measurements of the adsorption of proteins at interfaces. *J Colloid Interface Sci*. 1978;64(3):391-402. doi:10.1016/0021-9797(78)90382-X

197. Maget-Dana R. The monolayer technique: a potent tool for studying the interfacial properties of antimicrobial and membrane-lytic peptides and their interactions with lipid membranes. *Biochim Biophys Acta - Biomembr*. 1999;1462(1-2):109-140. doi:10.1016/S0005-2736(99)00203-5

198. Maget-Dana R, Leli□vre D, Brack A. Surface active properties of amphiphilic sequential isopeptides: Comparison between ?-helical and ?-sheet conformations. *Biopolymers*. 1999;49(5):415-423. doi:10.1002/(SICI)1097-0282(19990415)49:5<415::AID-BIP7>3.0.CO;2-J

199. Taylor JW. Peptide models of dynorphin A(1-17) incorporating minimally homologous substitutes for the potential amphiphilic .beta. strand in residues 7-15. *Biochemistry*. 1990;29(22):5364-5373. doi:10.1021/bi00474a023

200. MacRitchie F, Alexander A. Kinetics of adsorption of proteins at interfaces. Part II. The role of pressure barriers in adsorption. *J Colloid Sci*. 1963;18(5):458-463. doi:10.1016/0095-8522(63)90037-0

201. Hoernke M, Falenski JA, Schwieger C, Koksch B, Brezesinski G. Triggers for β-Sheet Formation at the Hydrophobic–Hydrophilic Interface: High Concentration, In-Plane Orientational Order, and Metal Ion Complexation. *Langmuir*. 2011;27(23):14218-14231. doi:10.1021/la203016z

202. Lad MD, Birembaut F, Matthew JM, Frazier RA, Green RJ. The adsorbed conformation of globular proteins at the air/water interface. *Phys Chem Chem Phys*. 2006;8(18):2179. doi:10.1039/b515934b

203. Yano YF, Arakawa E, Voegeli W, Kamezawa C, Matsushita T. Initial Conformation of Adsorbed Proteins at an Air–Water Interface. *J Phys Chem B*. 2018;122(17):4662-

4666. doi:10.1021/acs.jpcb.8b01039

204. Yang Y, Dicko C, Bain CD, et al. Behavior of silk protein at the air–water interface. *Soft Matter*. 2012;8(37):9705. doi:10.1039/c2sm26054a

205. Akram MS. Functionalised self-assembling polyglutamine fusion tags. 2012.

206. Rehman JU. Polyglutamine tagged functional fusion proteins. 2015.

207. Kubin RF, Fletcher AN. Fluorescence quantum yields of some rhodamine dyes. *J Lumin*. 1982;27(4):455-462. doi:10.1016/0022-2313(82)90045-X

208. Magde D, Rojas GE, Seybold PG. Solvent Dependence of the Fluorescence Lifetimes of Xanthene Dyes. *Photochem Photobiol*. 1999;70(5):737-744. doi:10.1111/j.1751-1097.1999.tb08277.x

209. Gunawardena J. Some lessons about models from Michaelis and Menten. Kellogg D, ed. *Mol Biol Cell*. 2012;23(4):517-519. doi:10.1091/mbc.e11-07-0643

210. Mortazavi M, Hosseinkhani S, Khajeh K, Ranjbar B, Emamzadeh AR. Spectroscopic and functional characterization of Lampyris turkestanicus luciferase: a comparative study. *Acta Biochim Biophys Sin (Shanghai)*. 2008;40(5):365-374. doi:10.1111/j.1745-7270.2008.00411.x

211. Wei Y, Thyparambil AA, Latour RA. Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochim Biophys Acta - Proteins Proteomics*. 2014;1844(12):2331-2337. doi:10.1016/j.bbapap.2014.10.001

212. Khurana R, Coleman C, Ionescu-Zanetti C, et al. Mechanism of thioflavin T binding to amyloid fibrils. *J Struct Biol*. 2005;151(3):229-238. doi:10.1016/j.jsb.2005.06.006

213. Http://isoelectric.org/index.html. Protein isoelectric point calculator.

214. Onodera O, Rose AD, Tsuji S, Vance JM, Strittmatter WJ, Burke JR. Toxicity of expanded polyglutamine-domain proteins in Escherichia coli. *FEBS Lett*. 1996;399(1-2):135-139. doi:10.1016/S0014-5793(96)01301-4

215. Bindels DS, Haarbosch L, van Weeren L, et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods*. 2017;14(1):53-56. doi:10.1038/nmeth.4074

216. Pandelieva A. Increasing the Quantum Yield of Red Fluorescent Proteins Using Rational Design. 2016. http://hdl.handle.net/10393/34272%0Ahttp://dx.doi.org/10.20381/ruor-5309.

217. Stepanenko O V., Verkhusha V V., Kazakov VI, et al. Comparative Studies on the Structure and Stability of Fluorescent Proteins EGFP, zFP506, mRFP1, “dimer2”, and DsRed1 †. *Biochemistry*. 2004;43(47):14913-14923. doi:10.1021/bi048725t

218. Peter M, Ameer-Beg SM, Hughes MKY, et al. Multiphoton-FLIM Quantification of the EGFP-mRFP1 FRET Pair for Localization of Membrane Receptor-Kinase Interactions. *Biophys J*. 2005;88(2):1224-1237. doi:10.1529/biophysj.104.050153

219. Zako T, Ayabe K, Aburatani T, et al. Luminescent and substrate binding activities of firefly luciferase N-terminal domain. *Biochim Biophys Acta - Proteins Proteomics*. 2003;1649(2):183-189. doi:10.1016/S1570-9639(03)00179-1

220. Marques SM, Esteves da Silva JCG. Firefly bioluminescence: A mechanistic approach of luciferase catalyzed reactions. *IUBMB Life*. 2009;61(1):6-17. doi:10.1002/iub.134

221. Borroto-Escuela DO, Flajolet M, Agnati LF, Greengard P, Fuxe K. Bioluminescence Resonance Energy Transfer Methods to Study G Protein-Coupled Receptor–Receptor Tyrosine Kinase Heteroreceptor Complexes. In: ; 2013:141-164. doi:10.1016/B978-0-12-408143-7.00008-6

222. Yang H, Yang S, Kong J, Dong A, Yu S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat Protoc*. 2015;10(3):382-396. doi:10.1038/nprot.2015.024

223. Barth A. Infrared spectroscopy of proteins. *Biochim Biophys Acta - Bioenerg*. 2007;1767(9):1073-1101. doi:10.1016/j.bbabio.2007.06.004

224. Adochitei A, Drochioiu G. Rapid characterization of peptide secondary structure by FT-IR spectroscopy. *Rev Roum Chim*. 2011;56(8):783-791.

225. Robetta. https://robetta.bakerlab.org. Accessed July 18, 2020.

226. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. 2004;32(Web Server):W526-W531. doi:10.1093/nar/gkh468

227. Ramachandran G, Udgaonkar JB. Understanding the Kinetic Roles of the Inducer Heparin and of Rod-like Protofibrils during Amyloid Fibril Formation by Tau Protein. *J Biol Chem*. 2011;286(45):38948-38959. doi:10.1074/jbc.M111.271874

228. Piatkevich KD, Verkhusha V V. Guide to Red Fluorescent Proteins and Biosensors for Flow Cytometry. In: ; 2011:431-461. doi:10.1016/B978-0-12-374912-3.00017-1

229. Damodaran S. Protein Stabilization of Emulsions and Foams. *J Food Sci*. 2006;70(3):R54-R66. doi:10.1111/j.1365-2621.2005.tb07150.x

230. Hoffmann JA, Hetru C. Insect defensins: inducible antibacterial peptides. *Immunol Today*. 1992;13(10):411-415. doi:10.1016/0167-5699(92)90092-L

231. Liao Z, Lampe JW, Ayyaswamy PS, Eckmann DM, Dmochowski IJ. Protein Assembly at the Air–Water Interface Studied by Fluorescence Microscopy. *Langmuir*. 2011;27(21):12775-12781. doi:10.1021/la203053g

232. Byler DM, Susi H. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers*. 1986;25(3):469-487. doi:10.1002/bip.360250307

233. Zhang K, Si FW, Duan HL, Wang J. Microstructures and mechanical properties of silks of silkworm and honeybee. *Acta Biomater*. 2010;6(6):2165-2171. doi:10.1016/j.actbio.2009.12.030

234. Liu Y, Shao Z, Vollrath F. Elasticity of Spider Silks. *Biomacromolecules*.

2008;9(7):1782-1786. doi:10.1021/bm7014174

235. Panwar P, Lamour G, Mackenzie NCW, et al. Changes in Structural-Mechanical Properties and Degradability of Collagen during Aging-associated Modifications. *J Biol Chem*. 2015;290(38):23291-23306. doi:10.1074/jbc.M115.644310

236. Bowen CH, Dai B, Sargent CJ, et al. Recombinant Spidroins Fully Replicate Primary Mechanical Properties of Natural Spider Silk. *Biomacromolecules*. 2018;19(9):3853-3860. doi:10.1021/acs.biomac.8b00980

237. Zhu Z, Kikuchi Y, Kojima K, et al. Mechanical Properties of Regenerated Bombyx mori Silk Fibers and Recombinant Silk Fibers Produced by Transgenic Silkworms. *J Biomater Sci Polym Ed*. 2010;21(3):395-411. doi:10.1163/156856209X423126

238. IRWIN GR. Studies in Large Plastic Flow and Fracture: With Special Emphasis on the Effects of Hydrostatic Pressure. P. W. Bridgman. New York-London: McGraw-Hill, 1952. 362 pp. $8.00. *Science (80-)*. 1952;115(2990):424-424. doi:10.1126/science.115.2990.424

239. Smith BL, Schäffer TE, Viani M, et al. Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature*. 1999;399(6738):761-763. doi:10.1038/21607

Appendix

A.1 DNA sequences

mcRFP

CATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTC
GGGATCTGTACGACGATGACGATAAGGATCCGATGGCCTCCTCCGAGGACGTCA
TCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACG
AGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACC
GCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTG
TCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCC
CCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAA
CTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGG
CGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTgtgACGGCCCCG
TAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGGATGTACCCCG
AGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTGAAGCTGAAGGACGGC
GGCCACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAGAAGCCCGTGCAG
CTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACATCACCTCCCACAACGAG
GACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCACTCCACCGGC
GCCTAA

mcRFP-Q77

CATATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTG
GACAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCCGATGGCCT
CCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCT
CCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACG
AGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCG
CCTGGGACATCCTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCA
CCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGG
GAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCC
TCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTC
CCCTgtgACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCG
AGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTG

AAGCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCACCTACATGGCC
AAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACATC
ACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGC
CGCCACTCCACCGGCGCCAAAGAATTCGAAGCCTATTTCGAAAAACAGCAGCAA
AAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC
AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
GCAGCAGCAGCAGCGGGACCTATCAGGATAA

Luc

ATGCATCACCATCACCATCACCCATGGGAAGACGCCAAAAACATAAAGAAAGGC
CCGGCACCACGCTATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAG
GCTATGAAGAGATACGCCCAGGTTCCTGGAACAATTGCTTTTACAGATGCACATA
TCGAGGTGAACATCACGTACGCGGAATACTTCGAAATGTCCGTTCGGTTGGCAG
AAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTG
AAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCA
GTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGAAC
ATTTCGCAGCCTACCGTAGTGTTTGTTTCCAAAAAGGGGTTGCAAAAAATTTTGA
ACGTGCAAAAAAAATTACCAATAATCCAGAAAATTATTATCATGGATTCTAAAA
CGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCC
CGGTTTTAATGAATACGATTTTAAACCAGAAAGCTTTGATCGTGACAAAACAATT
GCACTGATAATGAATTCCTCTGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTC
CGCATAGAACTGCCTGCGTCAGATTCTCGCACGCCAGAGATCCTATTTTTGGCAA
TCAAATCAAACCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTG
GAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTAT
AGATTTGAAGAAGAGCTGTTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTG
CGTTGCTAGTACCAACCCTATTTTCATTCTTCGCCAAAAGCACTCTGATTGACAA
ATACGATTTATCTAATTTACACGAAATTGCTTCTGGGGGCGCACCTCTTTCGAAA
GAAGTCGGGGAAGCGGTTGCAAAACGCTTCCATCTTCCAGGGATACGACAAGGA
TATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATA
AACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCT
GGATACCGGGAAAACGCTGGGCGTTAATCAGAGAGGCGAATTATGTGTCAGAGG
ACCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATT
GACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAA
CACTTCTTCATAGTTGACCGCTTGAAGTCTTTAATTAAATACAAAGGATATCAGG
TGGCCCCCGCTGAATTGGAATCGATATTGTTACAACACCCCAACATCCGCGACGC
CGGCGTGGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTT
GTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCC
AGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAA

GTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTC
ATAAAGGCCAAGAAGGGCGGAAAGTCCAAATTGTAA

Luc-Q77

ATGCATCACCATCACCATCACCCATGGGAAGACGCCAAAAACATAAAGAAAGGC
CCGGCACCACGCTATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAG
GCTATGAAGAGATACGCCCAGGTTCCTGGAACAATTGCTTTTACAGATGCACATA
TCGAGGTGAACATCACGTACGCGGAATACTTCGAAATGTCCGTTCGGTTGGCAG
AAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTG
AAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCA
GTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGAAC
ATTTCGCAGCCTACCGTAGTGTTTGTTTCCAAAAAGGGGTTGCAAAAAATTTTGA
ACGTGCAAAAAAAATTACCAATAATCCAGAAAATTATTATCATGGATTCTAAAA
CGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCC
CGGTTTTAATGAATACGATTTTAAACCAGAAAGCTTTGATCGTGACAAAACAATT
GCACTGATAATGAATTCCTCTGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTC
CGCATAGAACTGCCTGCGTCAGATTCTCGCACGCCAGAGATCCTATTTTTGGCAA
TCAAATCAAACCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTG
GAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTAT
AGATTTGAAGAAGAGCTGTTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTG
CGTTGCTAGTACCAACCCTATTTTCATTCTTCGCCAAAAGCACTCTGATTGACAA
ATACGATTTATCTAATTTACACGAAATTGCTTCTGGGGGCGCACCTCTTTCGAAA
GAAGTCGGGGAAGCGGTTGCAAAACGCTTCCATCTTCCAGGGATACGACAAGGA
TATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATA
AACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCT
GGATACCGGGAAAACGCTGGGCGTTAATCAGAGAGGCGAATTATGTGTCAGAGG
ACCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATT
GACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAA
CACTTCTTCATAGTTGACCGCTTGAAGTCTTTAATTAAATACAAAGGATATCAGG
TGGCCCCCGCTGAATTGGAATCGATATTGTTACAACACCCCAACATCCGCGACGC
CGGCGTGGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTT
GTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCC
AGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAA
GTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTC
ATAAAGGCCAAGAAGGGCGGAAAGTCCAAATTGCCAAGATTTAGATTTCGGGCT
AGCACTAGTGGATCCAAGCTTGAAAATCTGTATTTTCAGGGCGAATTCGAAGCCT
ATTTCGAAAAACAGCAGCAAAAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC

AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCGGGACCTATCAGGATAA

Luc-mcRFP

ATGCATCACCATCACCATCACCCATGGGAAGACGCCAAAAACATAAAGAAAGGC
CCGGCACCACGCTATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAG
GCTATGAAGAGATACGCCCAGGTTCCTGGAACAATTGCTTTTACAGATGCACATA
TCGAGGTGAACATCACGTACGCGGAATACTTCGAAATGTCCGTTCGGTTGGCAG
AAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTG
AAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCA
GTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGAAC
ATTTCGCAGCCTACCGTAGTGTTTGTTTCCAAAAAGGGGTTGCAAAAAATTTTGA
ACGTGCAAAAAAAATTACCAATAATCCAGAAAATTATTATCATGGATTCTAAAA
CGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCC
CGGTTTTAATGAATACGATTTTAAACCAGAAAGCTTTGATCGTGACAAAACAATT
GCACTGATAATGAATTCCTCTGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTC
CGCATAGAACTGCCTGCGTCAGATTCTCGCACGCCAGAGATCCTATTTTTGGCAA
TCAAATCAAACCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTG
GAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTAT
AGATTTGAAGAAGAGCTGTTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTG
CGTTGCTAGTACCAACCCTATTTTCATTCTTCGCCAAAAGCACTCTGATTGACAA
ATACGATTTATCTAATTTACACGAAATTGCTTCTGGGGGCGCACCTCTTTCGAAA
GAAGTCGGGGAAGCGGTTGCAAAACGCTTCCATCTTCCAGGGATACGACAAGGA
TATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATA
AACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCT
GGATACCGGGAAAACGCTGGGCGTTAATCAGAGAGGCGAATTATGTGTCAGAGG
ACCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATT
GACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAA
CACTTCTTCATAGTTGACCGCTTGAAGTCTTTAATTAAATACAAAGGATATCAGG
TGGCCCCCGCTGAATTGGAATCGATATTGTTACAACACCCCAACATCCGCGACGC
CGGCGTGGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTT
GTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCC
AGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAA
GTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTC
ATAAAGGCCAAGAAGGGCGGAAAGTCCAAATTGCCAAGATTTAGATTTCGGGAT
CCGATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCA
TGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGC
CGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCC

CTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCT
ACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGG
CTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGAC
CCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGG
CACCAACTTCCCCTGTGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGA
GGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAA
GATGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCA
CCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCA
AGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGC
GCGCCGAGGGCCGCCACTCCACCGGCGCCTAA

Luc-mcRFP-Q77

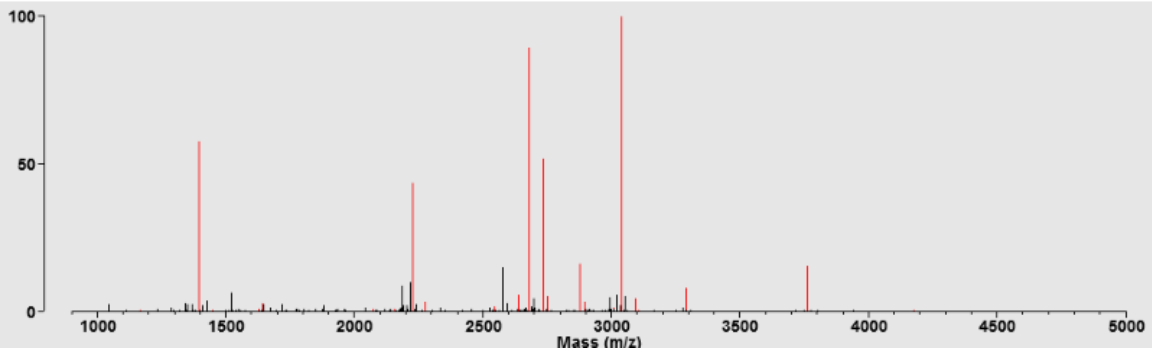
ATGCATCACCATCACCATCACCCATGGGAAGACGCCAAAAACATAAAGAAAGGC
CCGGCACCACGCTATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAG
GCTATGAAGAGATACGCCCAGGTTCCTGGAACAATTGCTTTTACAGATGCACATA
TCGAGGTGAACATCACGTACGCGGAATACTTCGAAATGTCCGTTCGGTTGGCAG
AAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTG
AAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCA
GTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGAAC
ATTTCGCAGCCTACCGTAGTGTTTGTTTCCAAAAAGGGGTTGCAAAAAATTTTGA
ACGTGCAAAAAAAATTACCAATAATCCAGAAAATTATTATCATGGATTCTAAAA
CGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCC
CGGTTTTAATGAATACGATTTTAAACCAGAAAGCTTTGATCGTGACAAAACAATT
GCACTGATAATGAATTCCTCTGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTC
CGCATAGAACTGCCTGCGTCAGATTCTCGCACGCCAGAGATCCTATTTTTGGCAA
TCAAATCAAACCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTG
GAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTAT
AGATTTGAAGAAGAGCTGTTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTG
CGTTGCTAGTACCAACCCTATTTTCATTCTTCGCCAAAAGCACTCTGATTGACAA
ATACGATTTATCTAATTTACACGAAATTGCTTCTGGGGGCGCACCTCTTTCGAAA
GAAGTCGGGGAAGCGGTTGCAAAACGCTTCCATCTTCCAGGGATACGACAAGGA
TATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATA
AACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCT
GGATACCGGGAAAACGCTGGGCGTTAATCAGAGAGGCGAATTATGTGTCAGAGG
ACCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATT
GACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAA
CACTTCTTCATAGTTGACCGCTTGAAGTCTTTAATTAAATACAAAGGATATCAGG
TGGCCCCCGCTGAATTGGAATCGATATTGTTACAACACCCCAACATCCGCGACGC
CGGCGTGGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTT
GTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCC

AGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAA
GTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTC
ATAAAGGCCAAGAAGGGCGGAAAGTCCAAATTGCCAAGATTTAGATTTCGGGAT
CCGATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGTGCGCA
TGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGC
CGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCC
CTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCCAGTACGGCTCCAAGGCCT
ACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGG
CTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGAC
CCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGG
CACCAACTTCCCCTGTGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGA
GGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAA
GATGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCA
CCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCA
AGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGC
GCGCCGAGGGCCGCCACTCCACCGGCGCCAAAGAATTCGAAGCCTATTTCGAAA
AACAGCAGCAAAAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC
AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCGGGACCTATCAGGATAACTCGAG

A.2 Protein Mass Spectrum for Sequence Identification

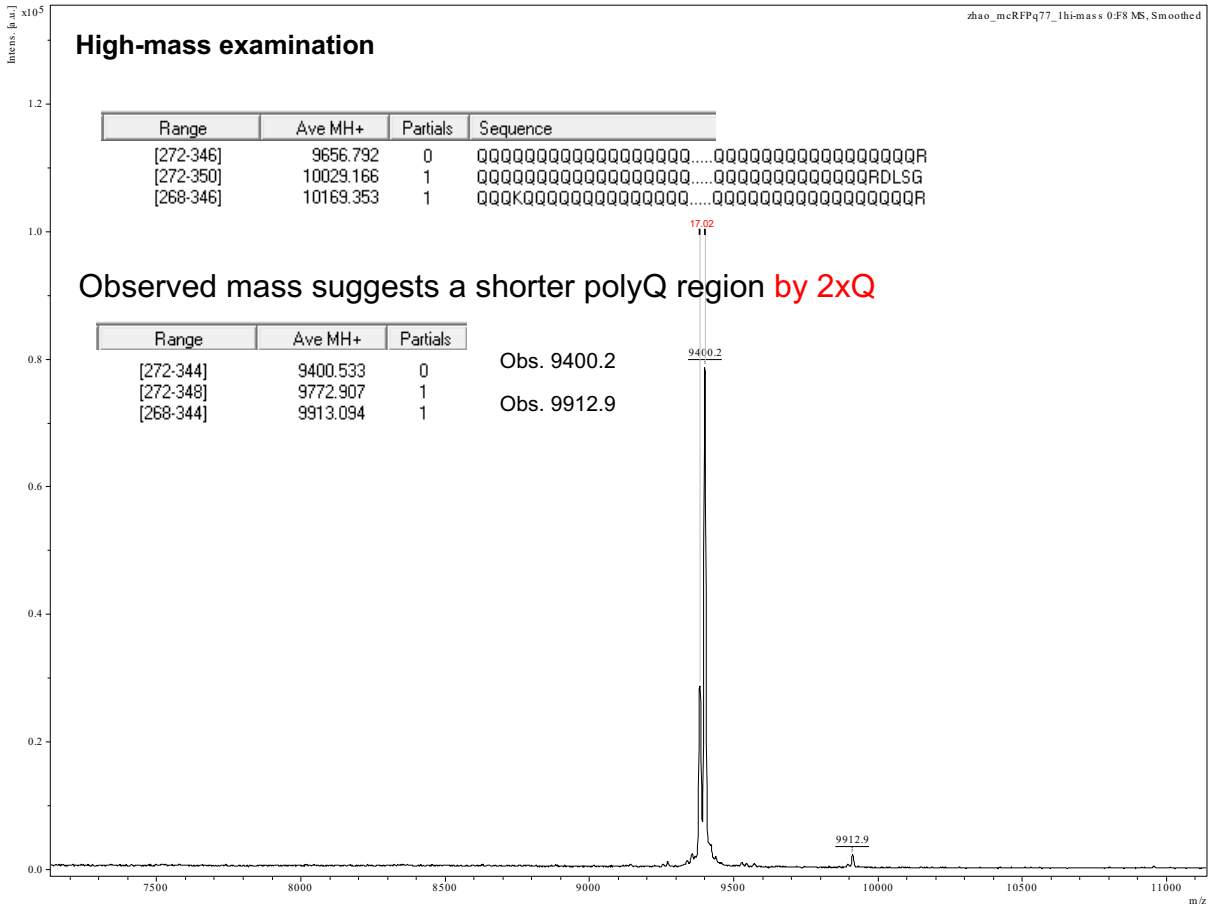
The matched peptides cover **63%** (220/350AA's) of the protein.

m/z Submitted	MH+ Matched	Intensity	Delta ppm	Modifications	Start	End	Missed Cleavages	Sequence
924.4850	924.4825	2361.0	2.66		118	125	0	(K)LSFPEGFK(W)
1062.4730	1062.4779	4769.0	-4.57		260	267	0	(K)EFEAYFEK(Q)
1100.6400	1100.6463	3198.0	-5.68		218	227	0	(K)KPVQLPGAYK(T)
1167.5080	1167.5099	8823.0	-1.63		173	182	0	(K)TMGWEASTER(M)
1395.7090	1395.7056	826037.0	2.46		118	128	1	(K)LSFPEGFKWER(V)
1450.7250	1450.7246	8384.0	0.246		183	195	1	(R)MYPEDGALKGEIK(M)
1629.8640	1629.8635	14302.0	0.295		104	117	1	(K)AYVKHPADIPDYLK(L)
1643.7740	1643.7700	42466.0	2.43		254	267	1	(R)HSTGAKEFEAYFEK(Q)
1875.9950	1875.9745	11215.0	10.9	1Oxidation	196	211	3	(K)MRLKLKDGGHYDAEVK(T)
2074.0670	2074.0644	13314.0	1.25		108	125	1	(K)HPADIPDYLKLSFPEGFK(W)
2156.0420	2156.0407	14770.0	0.599		254	271	2	(R)HSTGAKEFEAYFEKQQQK(Q)
2225.0380	2225.0357	623335.0	1.04		232	249	0	(K)LDITSHNEDYTIVEQYER(A)
2277.9680	2277.9672	47244.0	0.351		3	23	0	(R)GSHHHHHHGMASMTGGQQMGR(D)
2545.2840	2545.2874	27135.0	-1.34		108	128	2	(K)HPADIPDYLKLSFPEGFKWER(V)
2638.2570	2638.2380	83356.0	7.22		232	253	1	(K)LDITSHNEDYTIVEQYERAEGR(H)
2682.2890	2682.2893	1281303.0	-0.123		228	249	1	(K)TDIKLDITSHNEDYTIVEQYER(A)
2734.1840	2734.1859	740929.0	-0.678		24	46	2	(R)DLYDDDDKDPMASSEDVIKEFMR(F)
2750.1780	2750.1808	75244.0	-1.01	1Oxidation	24	46	2	(R)DLYDDDDKDPMASSEDVIKEFMR(F)
2878.3370	2878.3451	232775.0	-2.83		129	154	0	(R)VMNFEDGGVVTVTQDSSLQDGEFIYK(V)
2894.3320	2894.3400	47598.0	-2.78	1Oxidation	129	154	0	(R)VMNFEDGGVVTVTQDSSLQDGEFIYK(V)
3006.5480	3006.5512	18308.0	-1.08		104	128	3	(K)AYVKHPADIPDYLKLSFPEGFKWER(V)
3038.3430	3038.3432	1427840.0	-0.0770		51	78	0	(R)MEGSVNGHEFEIEGEGEGRPYEGTQTAK(L)
3095.4890	3095.4916	64452.0	-0.844		228	253	2	(K)TDIKLDITSHNEDYTIVEQYERAEGR(H)
3105.5040	3105.5085	9170.0	-1.45		129	156	1	(R)VMNFEDGGVVTVTQDSSLQDGEFIYKVK(L)
3293.5160	3293.5128	117043.0	0.984		49	78	1	(K)VRMEGSVNGHEFEIEGEGEGRPYEGTQTAK(L)
3763.9150	3763.9177	221438.0	-0.730		218	249	2	(K)KPVQLPGAYKTDIKLDITSHNEDYTIVEQYER(A)
4177.1150	4177.1200	9313.0	-1.20		218	253	3	(K)KPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGR(H)



Name: **mcRFP-Q77**

```
1   MRGSHHHHHH GMASMTGGQQ MGRDLYDDDD KDPMASSEDV IKEFMRFKVR MEGSVNGHEF EIEGEGEGRP YEGTQTAKLK
81  VTKGGPLPFA WDILSPQFQY GSKAYVKHPA DIPDYLKLSF PEGFKWERVM NFEDGGVVTV TQDSSLQDGE FIYKVKLRGT
161 NFPCDGPVMQ KKTMGWEAST ERMYPEDGAL KGEIKMRLKL KDGGHYDAEV KTTYMAKKPV QLPGAYKTDI KLDITSHNED
241 YTIVEQYERA EGRHSTGAKE FEAYFEKQQQ KQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ
321 QQQQQQQQQQ QQQQQQQQQQ QQQQQRDLSG
```


High-mass examination
Range | Ave MH+ | Partials | Sequence
[272-346] 9656.792 0 QQQQQQQQQQQQQQQQQQ.....QQQQQQQQQQQQQQQQQR
[272-350] 10029.166 1 QQQQQQQQQQQQQQQQQQ.....QQQQQQQQQQQQQRDLSG
[268-346] 10169.353 1 QQQKQQQQQQQQQQQQQQ.....QQQQQQQQQQQQQQQQQR
Observed mass suggests a shorter polyQ region by 2xQ
Range | Ave MH+ | Partials
[272-344] 9400.533 0
[272-348] 9772.907 1
[268-344] 9913.094 1
Obs. 9400.2
Obs. 9912.9
17.02
9400.2
9912.9
zhao_mcRFPq77_1hi-mass 0:F8 MS, Smoothed
Intens. [a.u.]
x10^5
m/z

A.3 Aggregation Images

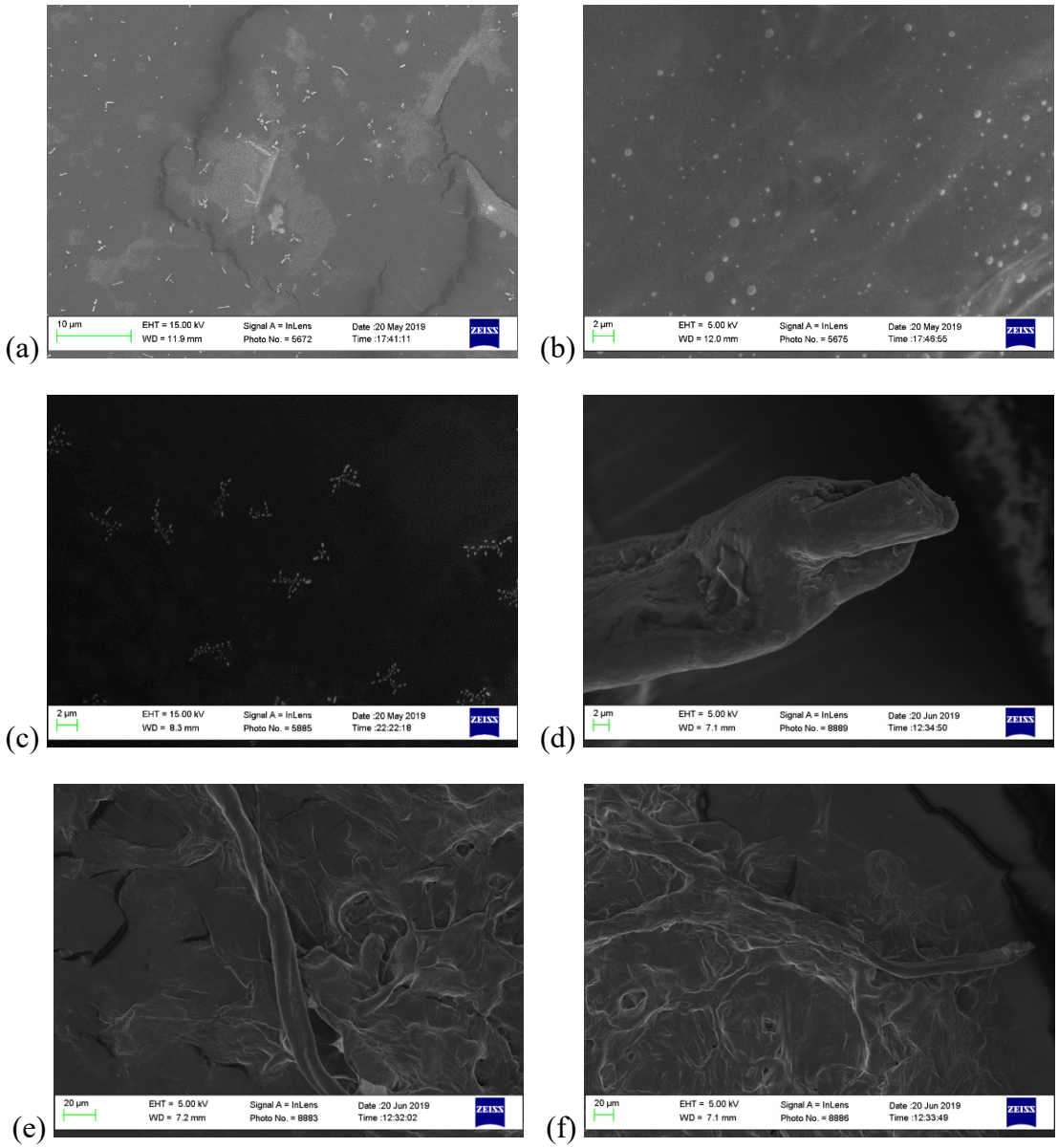


Figure A3.1– SEM images of mcRFP-Q77 self-assembled after 1-day incubation at (a) pH 3; (b) pH 5; (c) pH 8; and (d)-(f) after 2 weeks of incubation at pH 8 at 5 µM, room temperature.

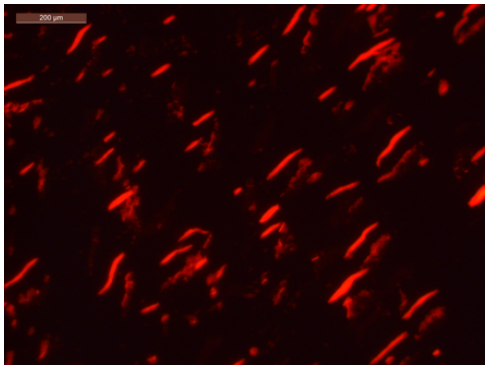
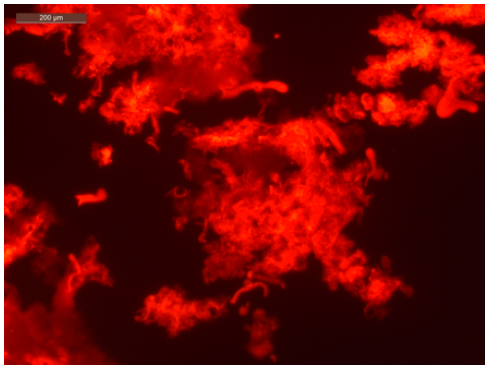
(a)  (b) 

Figure A3.2– Fluorescent images of mcRFP-Q77 aggregates after (a) 1 week and (b) 2 weeks of assembly at room temperature at a concentration of 5 μM, pH 8. (Scale bar: 200 μm)

A.4 mcRFP and mcRFP-Q77 at AWI

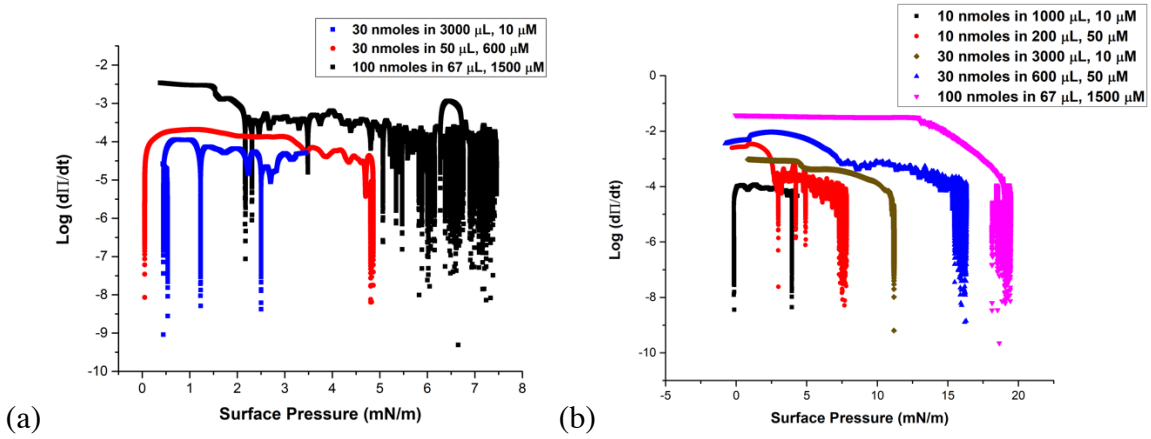


Figure A4.3– The assembly kinetics profile (log(dΠ/dt)-Π plot) at the AWI interface of (a) mcRFP and (b) mcRFP-Q77 in different molecular quantity at different initial concentration before spread on the interface.

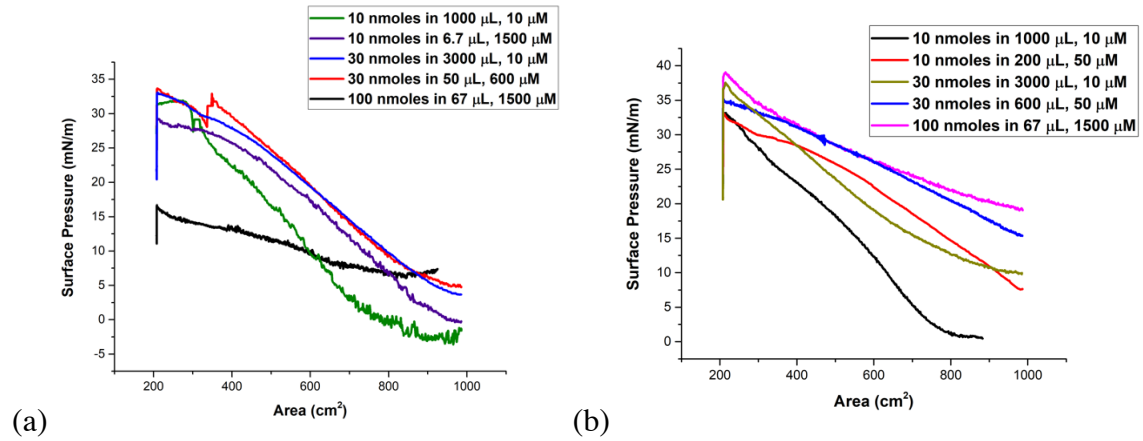
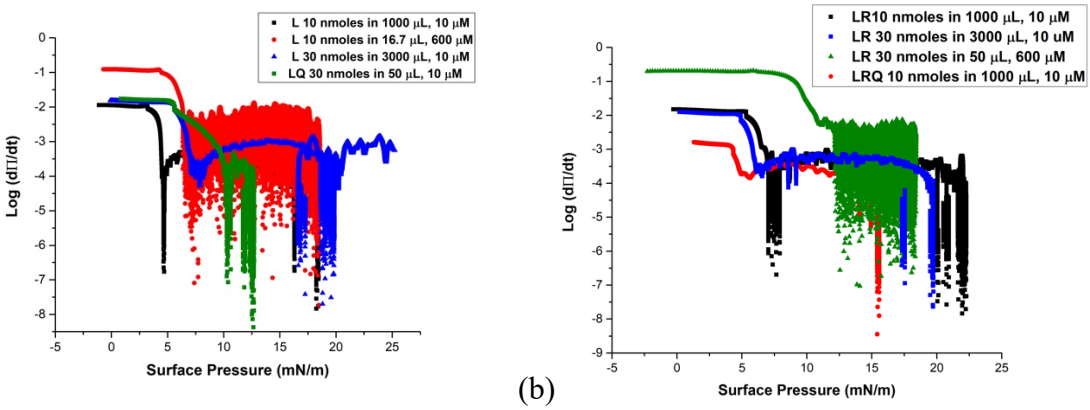


Figure A4.4– The compression profile (Π -A_0 plot, A_0 is the area of water surface within the two barriers) at AWI of (a) mcRFP and (b) mcRFP-Q77 in different molecular quantity at different initial concentration before spread on the interface, under a compression speed of 100 cm^2/min with two barriers.

Table A4.1 – The molecular footprint during the compression of the interface of mcRFP at different molecular quantity and different initial concentration.

	transition from G to L (Å²/molecule)		transition from LE to LC (Å²/molecule)		transition from L to S (Å²/molecule)		monolayer collapse (Å²/molecule)		monolayer collapse (mN/m)	
	R	RQ	R	RQ	R	RQ	R	RQ	R	RQ
10 nmoles, 10 μM	1235	1289	784	860	668		466		32	
10 nmoles, 1500 μM	1527	1964	807	667						
30 nmoles, 10 μM	556	562	249	350						
30 nmoles, 50 μM				256						
30 nmoles, 600 μM				543						
100 nmoles, 1500 μM					109	155				

A.5 Luc Constructs at AWI



(a) (b)

Figure A5.5– The assembly kinetics profile (log(dΠ/dt)-Π plot) at the AWI interface of (a) Luc and Luc-Q77, (b) Luc-mcRFP and Luc-mcRFP-Q77 in different molecular quantity at different initial concentration before spread on the interface.

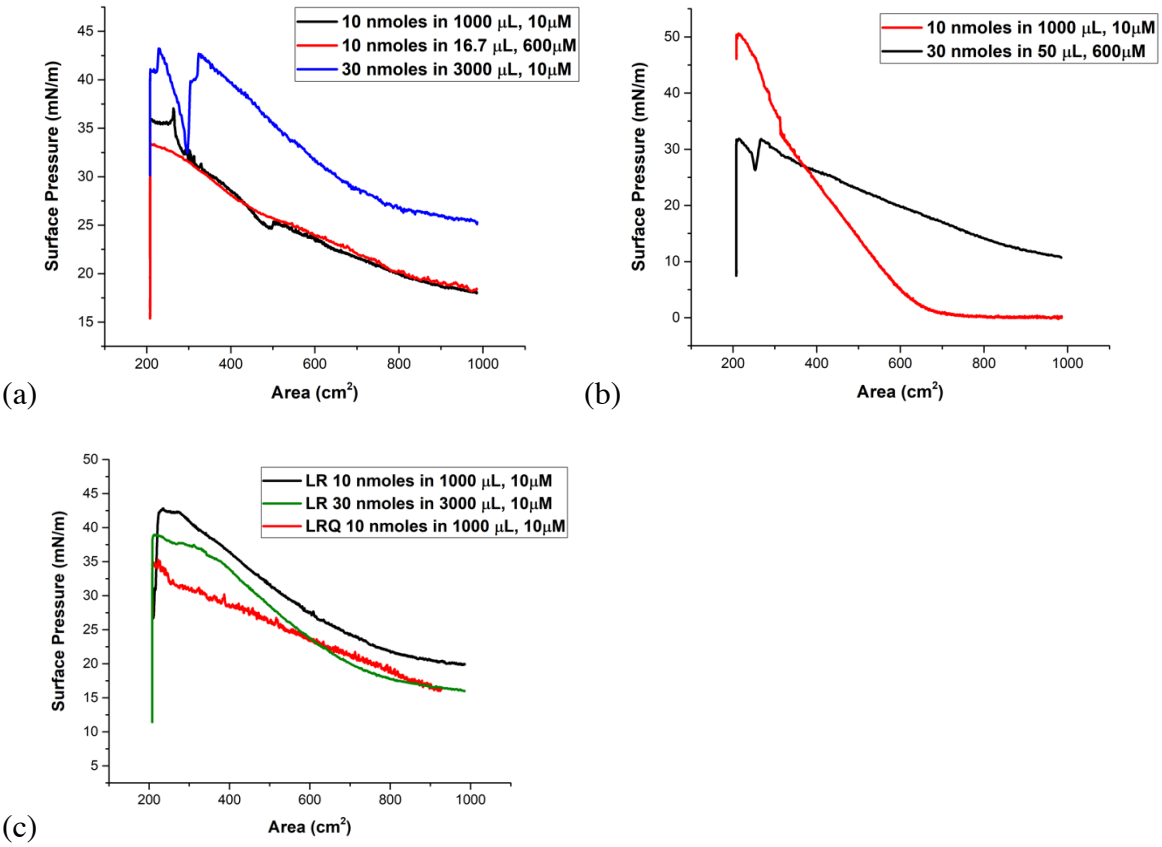


Figure A5.6– The compression profile (Π -A_0 plot, A_0 is the area of water surface within the two barriers) at AWI of (a) Luc, (b) Luc-Q77, (c) Luc-mcRFP and Luc-mcRFP-Q77 in different molecular quantity at different initial concentration before spread on the interface, under a compression speed of 100 cm2/min with two barriers.

Table A5.2 – The molecular footprint during the compression of the interface of Luc and Luc-Q77 at different molecular quantity and different initial concentration.

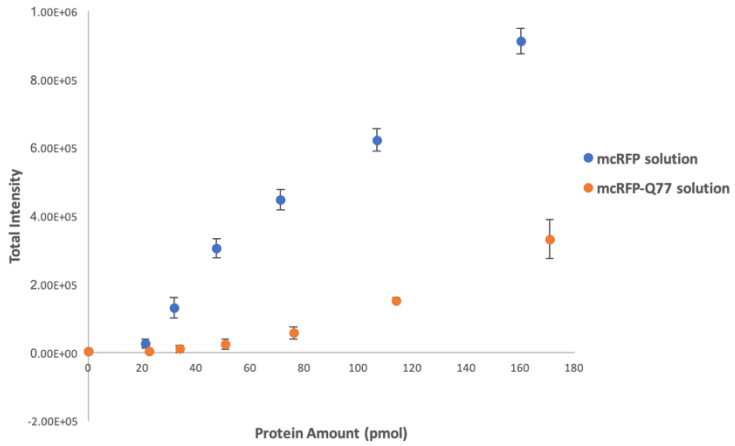
	transition from G to L (Å²/molecule)		transition from LE to LC (Å²/molecule)		transition from L to S (Å²/molecule)		monolayer collapse (Å²/molecule)		monolayer collapse (mN/m)	
	L	LQ	L	LQ	L	LQ	L	LQ	L	LQ
10 nmoles, 10 μM		1082		709	633		438		37	
10 nmoles, 600 μM	2990		989							
30 nmoles, 10 μM					292		140		43	
30 nmoles, 600 μM						222		125		32

Table A5.3 – The molecular footprint during the compression of the interface of Luc-mcRFP and Luc-mcRFP-Q77 at different molecular quantity and different initial concentration.

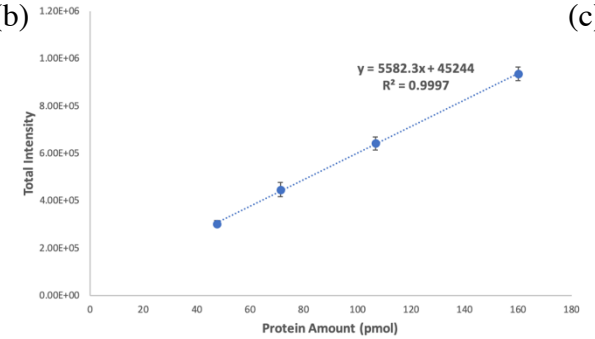
	transition from G to L (Å²/molecule)		transition from LE to LC (Å²/molecule)		transition from L to S (Å²/molecule)		monolayer collapse (Å²/molecule)		monolayer collapse (mN/m)	
	LR	LRQ	LR	LRQ	LR	LRQ	LR	LRQ	LR	LRQ
10 nmoles, 10 μM			1751	988			500		42	
30 nmoles, 10 μM			189							

A.6 Calibration Curves

(a)



(b)



(c)

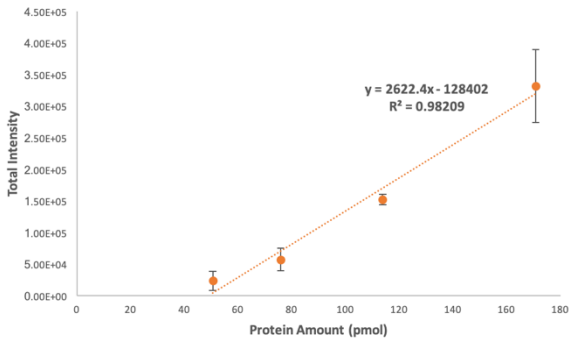


Figure A6.7– Fluorescent assay of (a) full range of mcRFP and mcRFP-Q77 native solution, and linear range of (b) native state mcRFP solution and (c) native state mcRFP-Q77 solution under BiognostiX reader.

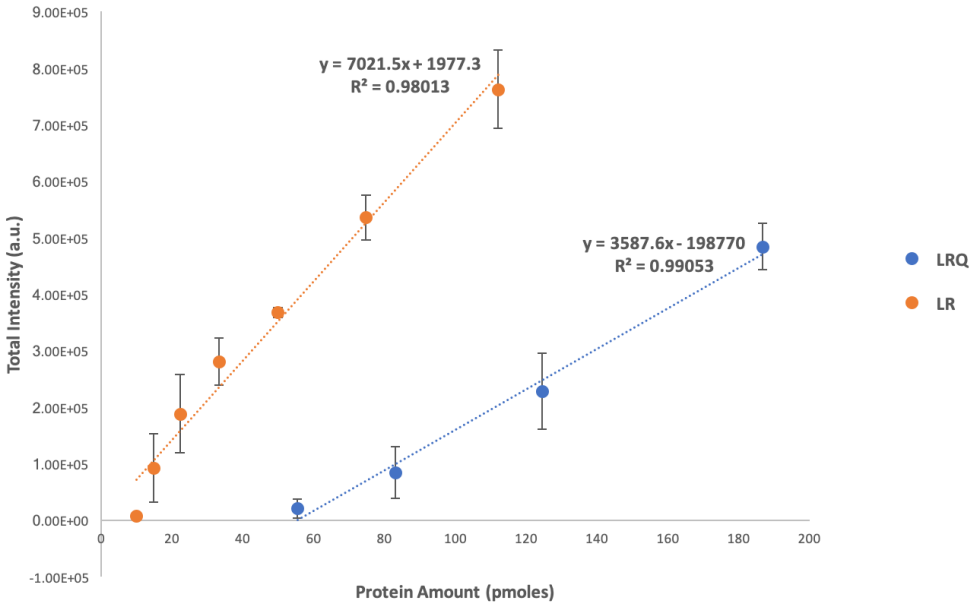


Figure A6.8– Fluorescent assay comparing the linear range of native state Luc-mcRFP and Luc-mcRFP-Q77 solution under BiognostiX reader.

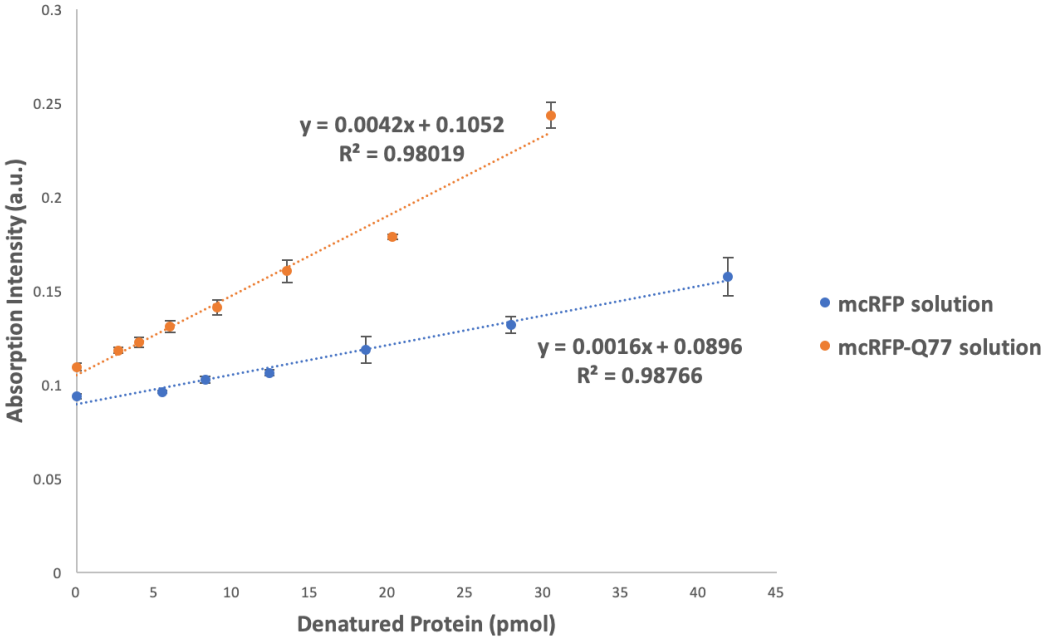


Figure A6.9– BCA assay comparing urea denatured mcRFP solution, and mcRFP-Q77 solution.

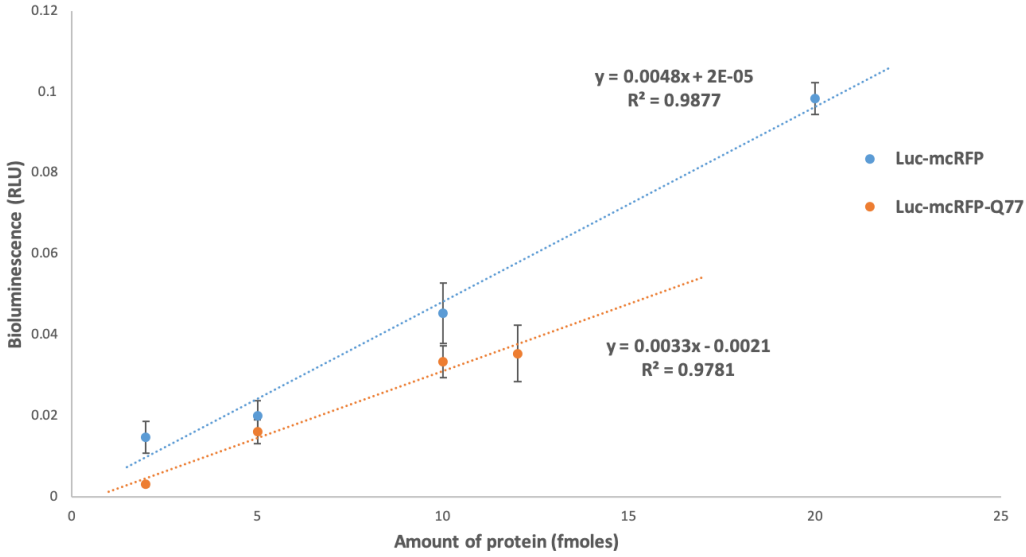


Figure A6.10– Bioluminescent activity assay of native state Luc-mcRFP and Luc-mcRFP-Q77 in solution.